Universidad Autónoma de Madrid
Escuela Politécnica Superior

Trabajo de fin de Grado

RECOMENDANDO ACCIONES BASADAS EN LEARNING ANALYTICS PARA MEJORAR EL RENDIMIENTO ACADÉMICO

Ingeniería de Informática

Miguel Ángel Álvarez Rodríguez
Junio 2018

RECOMENDANDO ACCIONES BASADAS EN LEARNING ANALYTICS PARA MEJORAR EL RENDIMIENTO ACADÉMICO

AUTOR: Miguel Ángel Álvarez Rodríguez
TUTOR: Estrella Pulido Cañabate

Escuela Politécnica Superior
Universidad Autónoma de Madrid
Junio 2018

Resumen

Resumen

La finalidad de este Trabajo de Fin de Grado es el desarrollo de las herramientas necesarias para acercar las nuevas tecnologías a la educación. Estas herramientas servirán para el análisis de cursos online, en concreto los denominados MOOCs. Su uso está destinado a la docencia, para facilitar el estudio y análisis de los datos de forma sencilla. Gracias a estas herramientas resultará mucho más fácil detectar problemas durante el curso y conocer las causas y soluciones.

Este proyecto tiene como objetivo el uso aplicado en la educación del aprendizaje automático y effective learning analytics. Gracias a las herramientas desarrolladas los docentes pueden visualizar de forma sencilla los resultados del análisis. Lo que les ayuda a aplicar las acciones necesarias para solucionar los problemas encontrados durante el curso o realizar mejoras en él.

Estas herramientas extraen conclusiones sobre el comportamiento de los estudiantes mediante el análisis de su interacción con los vídeos y ejercicios del curso online. Gracias a este análisis, se puede observar cuántas veces se ve un vídeo, cuantos usuarios lo ven, qué partes se ven más, y qué zonas son las que más interacción tienen con el usuario. También relacionan los vídeos con los problemas mediante el uso de algoritmos de predicción. Esta relación permite identificar el temario común entre vídeos y problemas. Gracias a esto, se puede conocer cuántas preguntas tiene asociadas cada vídeo, si en el vídeo se explica correctamente el temario, si se detectan más errores que aciertos tras ver el vídeo, etc...

Además, este proyecto abre las puertas a un nuevo modelo educativo que permita realizar intervenciones en tiempo real a partir de los resultados obtenidos por los estudiantes hasta el momento. Estas intervenciones irán encaminadas a reducir el abandono y a mejorar los resultados.

Por último, a modo de resumen, este TFG supone el desarrollo de una serie de herramientas que permitan a los docentes mejorar el contenido de sus cursos de forma rápida y visual, analizando los vídeos y problemas publicados en el curso sobre el que se realice el estudio.

Palabras Clave

Aprendizaje automático, Effective Learning Analytics, árboles de decisión, educación, vídeos, problemas, MOOC, edX, SPOCs

Abstract

The purpose of this Final Degree Project is the development of the tools required to bring new technologies to education. These tools will be used to analyze online courses, specifically the so-called MOOCs. Its use is intended for teaching, to facilitate the study and analysis of data in a simple way. Thanks to these tools it will be much easier to detect problems during the course and know the causes and solutions.

The objective of this project is the applied use of machine learning in education and effective learning analytics. The result obtained thanks to the developed tools allow teachers to visualize the results in a simple way, being able to apply the necessary actions to solve the problems or improve some material in the course.

These tools are based on the behavior of students during the course, to perform the analysis of their interaction with the videos and exercises done in the online course. Thanks to them, it is possible to see how many times a video is viewed, how many users see it, which parts are more visualized more often, and which areas have the highest number of interactions with the user. They also relate the videos to the problems through the use of prediction algorithms. This relationship allows to identify the common contents between videos and problems. Thanks to this, It is possible to know how many questions are about each video, whether concepts are correctly explained in the video, if more error or more correct answers are detected after seeing the video, etc.

In addition, this project opens the doors to a new model educational that allow the implementation of interventions in real time based on the results obtained by student so far. These interventions would head towards the reduction of dropout and the improvement of academic results.

Finally, and as a summary, this TFG involves the development of a series of tools that allow teachers to improve the contents of their courses quickly and visually, by analyzing the videos and problems published in course being studied.

Key words

Machine Learning, Effective Learning Analytics, Decision Tree, Education, Videos, Problems TFG, MOOC, edX, SPOCs

Agradecimientos

Quiero empezar agradeciendo la ayuda de mi tutor, Estrella Pulido, por ofrecerme la posibilidad de desarrollar mi Trabajo de Fin de Grado en este campo y formar parte del equipo de la Cátedra UAM/IBM. También quiero incluir a Gonzalo Martínez, junto a Estrella, por ayudarme durante el desarrollo y visualización de los resultados, por darme feedback constante y proponerme ideas cuando no conseguíamos avanzar.

Quiero mencionar a mi compañero de laboratorio Ángel Pérez por las recomendaciones que me hacía cuando desarrollaba el TFG. También a mi otro compañero Miguel Ángel González-Gallego por su módulo de Preprocesamiento que nos ha facilitado el tratamiento inicial de los datos.

Finalmente quiero agradecer a mi familia por ayudarme a llegar hasta aquí y a mis amigos por interesarse en el TFG y todas sus aportaciones.

Índice general

Índice de figuras

1 Introducción

La educación ha sufrido numerosos cambios durante las últimas décadas, y en los últimos años la tecnología ha sido la responsable. En este TFG intentamos aportar en el acercamiento de la tecnología a la educación.

1.1. Motivación

La gran mayoría de las universidades españolas utilizan, actualmente, plataformas online en la docencia. Además la ingeniería informática permite estrechar la relación entre educación y nuevas tecnologías. Motivos por los cuales se ha fomentado, en los últimos años, el estudio y análisis del impacto de la tecnología en la educación.

Y cada vez su uso es mayor, no solo se utilizan para facilitar la disponibilidad del material utilizado, sino que están dando soporte a servicios como la calificación de las entregas, mensajes entre alumno y profesor, realización de exámenes, control de asistencia, etc... Estas plataformas han pasado de ser herramientas de apoyo a construir la base en muchas universidades. Todo el material del que dispone un alumno se limita a la plataforma online y los apuntes tomados en clase. En algunas asignaturas incluso podemos encontrar que no se toman apuntes, porque las clases son online o estos ya están en la plataforma.

Cada vez existen más herramientas que facilitan el uso de estas plataformas y son capaces de extraer muchos datos y estadísticas. Pero uno de los mayores problemas es la utilidad de estos datos. Podemos conocer cuándo un alumno se conecta, qué material ha visto y mucha otra información que sino se analizan termina sin utilidad.

Este es el principal problema de las herramientas actuales, son capaces de generan grandes cantidades de datos, tantas, que es imposible analizaras manualmente. Si bien, todos estos datos son bastante útiles en muchas ocasiones, resulta casi imposible para un docente realizar el estudio de todos los cursos y asignaturas por limitaciones de tiempo. Por lo que al final este estudio solo se realiza finalizado el curso, y pocas veces se obtienen resultados que supongan cambios en el siguiente curso.

1.2. Objetivos y enfoque

Ya existen numerosas herramientas para facilitar las tareas académicas del alumno. Además este solo tendría que analizar aquellos datos que le afecten personalmente. Por el contrario, el objetivo de este TFG es centrase en ayudar al profesorado y proporcionará las herramientas necesarias para mejorar desde la experiencia del alumnos hasta el material ofertado.

Uno de los grandes problemas en la educación es el elevado número de alumnos. Las aulas son cada vez más grandes y no es posible dedicar el tiempo suficiente para atender a cada alumno. Además, la tendencia es que la media de alumnos por curso aumente. Razón por la que los cursos online con miles de alumnos han crecido en los últimos años. Desarrollar una herramienta que automatice las tareas del profesorado y ayude en la detección de problemas en el curso es necesario.

Dentro de los distintos ámbitos de estudio nos centraremos en el uso de plataformas para impartir cursos MOOC (Massive Open Online Courses). Estos se centran en clases o explicaciones mediante vídeos y la utilización de ejercicios para comprobar los conocimientos adquiridos.

Nuestro objetivo es diseñar una herramienta que facilite la visualización e interpretación de los datos obtenidos. Esta herramienta permitirán ver cuántas veces es reproducido cada segundo del vídeo, qué partes del vídeo son más visualizadas y cuántos alumnos visualizan cada segundo de un vídeo, al menos una vez. Además relacionará los vídeos y problemas de manera que nos permita conocer: qué vídeos son más importantes, qué problemas son más difíciles y qué partes del vídeo son más visualizadas por los alumnos que aciertan frente aquellos que fallan. Esto ayudará a reforzar los vídeos difíciles, o confusos, a conocer qué materias necesitan ampliarse o cómo diseñar las preguntas del examen final.

Este es el objetivo principal, analizar los vídeos y conocer por qué un vídeo tiene algunas partes con muchas visualizaciones y otras con menos, y si esta información es relevante. Es decir, permitir analizar los vídeos con la información de los ejercicios, conocer cómo de repartido está el temario en un vídeo, si hay relación directa entre un vídeo y un problema, qué supone esta relación, qué partes del vídeo ayudan a acertar y cuáles no, permitir comparar el acierto y fallo de distintos problemas con las visualizaciones de un vídeo, si dos ejercicios son muy similares, etc...

En definitiva el objetivo es desarrollar una herramienta que permita analizar los vídeos y ejercicios, que sea capaz de determinar los pros y contras de cada vídeo y de manera sencilla y visual. Pero lo más importante es que los resultados se puedan usar para mejorar el contenido, ayudar a los alumnos y facilitar las tareas docentes.

1.3. Estructura del documento

Además de este capítulo de introducción, esta memoria se divide en otros 4 capítulos, que son los siguientes:

Capítulo 2 - Estado del arte

En este capítulo analizaremos qué herramientas similares a lo propuesto existen actualmente y que herramientas hemos utilizado.

Capítulo 3 - Preprocesamiento de los datos

En este capítulo detallaremos los datos iniciales y el formato, qué técnicas se han usado para el procesado inicial de los datos, qué modificaciones se han realizado y cómo. Finalmente mostraremos información general del curso.

Capítulo 4 - Análisis de los datos

Este capítulo es el más importante, pues en él, muestran los resultados obtenidos utilizando la herramienta desarrollada. Se describen los resultados obtenidos, su interpretación y utilidades.

Capitulo 5 - Conclusiones En este capítulo se presentan las conclusiones de trabajo y posibles líneas futuras de desarrollo.

2

Estado del arte

2.1. Historia, nacimiento y evolución.

Las nuevas tecnologías han experimentado un gran crecimiento durante la última década. Estas nuevas herramientas se han ido incorporado al sector educativo. La educación tradicional está siendo remplazada, y cada vez es más común el uso de material audiovisual, la utilización de plataformas online o los cursos por Internet. Tras la llegada de estas nuevas herramientas y metodologías, ha surgido la necesidad de su análisis y estudio.

El objetivo de estas nuevas tecnologías no es sustituir el sistema actual sino proporcionar las herramientas para mejorarlo. Por ejemplo las plataformas online como Moodle se centran en facilitar el acceso al material didáctico. Los cursos MOOC (Massive open online course) el acceso a cursos online [1] . La mayoría de herramientas audiovisuales se centran en hacer interesante las clases para el alumno. En particular, los vídeos han ido adquiriendo gran importancia, pues permiten al estudiante libertad y visualizar el contenido repetidas veces.

Aunque en muchas ocasiones las mejoras parecen claras, es necesario un estudio que corrobore los resultados. Actualmente la gran mayoría de estudios se centran en comparar los resultados con y sin el uso de estas herramientas [2] [3]. Dado que, generalmente, la conclusión es que el buen uso de estas herramientas mejora la calidad de la enseñanza, en los últimos años los estudios se han centrado en el uso de las distintas herramientas. Gran parte de los estudios se han centrado en determinar qué metodologías son las más adecuadas y en qué escenario aplicarlas [1]. Dónde y cómo aplicarla es muy importante, pues un buen contexto facilitará el análisis y estudio de los resultados [4].

En la última década el aprendizaje automático, las redes neuronales y la inteligencia artificial han adquirido gran importancia. Su aplicación se ha llevado a cabo en números sectores, entre ellos en la educación. Sus aplicaciones van desde predecir notas [5] a recomendaciones en el diseño de exámenes. Pero la gran mayoría, sobre todo en los últimos años, se han centrado en el aprendizaje automático [6].

Algunos estudios se centran en el estudio “a posteriori” de los resultados académicos. Obteniendo estadísticas, tendencias y prediciendo futuros resultados. Otros tienen el objetivo de a partir de los resultados del análisis influir en las decisiones y resultados, es lo que se denomina “Effective Machine learning” [7]. Algunos tienen intervenciones tempranas [8], los resultados

permiten conocer una estimación final de la nota y ayudar a los alumnos con bajo rendimiento académico o evitar el abandono. Otros se analizan tras finalizar el curso y se centran en mejorar el siguiente. Nuestra propuesta va en esta línea, con la diferencia de que las herramientas desarrolladas permiten una ejecución temprana pudiendo realizar intervenciones en el curso actual, al menos antes del examen final.

2.2. Herramientas

La implementación se ha realizado en Python 3.6. Para hacer más visual el entorno de ejecución se han utilizado los notebooks de la herramienta Jupyter. Esta permite ejecutar distintas partes del código, de manera que la ejecución está modulada, y permite, de forma conjunta, el uso de texto, donde se indica qué función realiza cada apartado dentro de un mismo notebook. A continuación se describen las librerías utilizadas durante el proyecto.

2.2.1. Plotly

Plotly es una librería para la generación de gráficos interactivos. Cuenta con soporte online y tiene gran variedad de herramientas para modificar, tras la ejecución, los datos, la visualización o la representación. El servicio online tiene restricciones según la suscripción activa, de modo que permite la generación de graficas offline. Estas gráficas cuentan con casi todas las herramientas del modo online y se generan en ficheros HTML que se pueden abrir en un navegador y simulan el uso online.

La gran ventaja de esta librería es la edición tras la generación. Una vez obtenidas las gráficas podemos modificarlas, cambiando sus dimensiones sin preocuparnos por la resolución, o modificando la escala y el espacio mostrado. Sin embargo, la utilidad más destacada es la selección de los datos a visualizar. En una única grafica podemos almacenar todos los valores para su posterior análisis, y aunque inicialmente sea demasiada información se pueden ocultar aquellos datos que no son necesarios o incluso mostrar solo los que nos interesen.

2.2.2. Scikit-learn

La librería Scikit-learn es la más usada en Python cuando se trata de Machine Learning. En este TFG se utiliza como para analizar la relación entre vídeos y problemas. La predicción de los resultados a partir de los vídeos nos permite obtener la relación que más tarde usaremos para mostrar los resultados analizados.

2.2.3. Pandas

Los ficheros usados cuentan con miles de líneas y cada una de estas con cientos de eventos. Debido a esta gran cantidad de datos es necesario una herramienta capaz de soportar estos tamaños. Mediante el uso de dataframes la librería Pandas nos permite operar sobre todos los datos de forma sencilla. Su principal ventaja es la aplicación de funciones a todos los datos de forma abstracta y la posibilidad de obtener valores como medias, desviaciones, máximos y mínimos de forma sencilla.

2.2.4. Otras librerías

Además de las librerías mencionadas anteriormente hemos usado:

1. Numpy, librería para computación científica y manejo de matrices en Python.
2. Datetime, librería para gestión de fechas y tiempo.
3. Math, librería matemática.
4. Re, librería de expresiones regulares.

2.3. Uso del programa y módulos

Para la obtención de todos los resultados son necesarias tres fases, además de una previa para la obtención del fichero inicial. Cada una de estas fases está separada en un notebook distinto de Jupyter.

La fase previa consiste en extraer los datos del curso analizado, que se encuentran en el log en formato plano mediante las herramientas desarrolladas en "Predicción y análisis de interacciones de usuarios en plataformas de enseñanza online"[9]. De este modo se extraen los datos y se guarda en un fichero con el formato JSON descrito en el capítulo 3.

La primera fase consiste en la obtención de las características generales del curso. Durante esta fase se extraen los datos que se usarán en el resto del proyecto por lo que es necesario su ejecución previa. También se generan nuevos a partir de los existentes, en particular lo que denominamos secuencias de visualización, que se usarán el resto de los módulos. Como se descubre en el capítulo 3, en esta fase se obtendrán resultados como la tasa de aciertos en los problemas y el número de usuarios en función del número de eventos realizados.

La segunda fase consiste en obtener información sobre los vídeos, como las visualizaciones totales de cada uno de los vídeos, el número de usuarios que visualizan cada segundo y el avance temporal de los eventos durante la reproducción de un vídeo. Los resultados de este análisis se detallan en profundidad en el capítulo 4.

La última fase consiste en analizar la relación entre problemas y vídeo. En esta fase se realizan los cálculos necesarios para determinar esta relación. A partir de esta información se generan distintas gráficas que se detallan en el capítulo 4 y permiten analizar en mayor profundidad cada vídeo y determinar sus problemas y soluciones.

3

Preprocesamiento de los Datos

En este capítulo hablaremos de las modificaciones realizadas sobre los datos originales y se analizará que efectos tienen. Las modificaciones realizadas incluyen filtrado de los datos o generación de nueva información partiendo de la anterior. Se explicarán las ventajas y desventajas de cada cambio, los resultados obtenidos y que objetivo buscan.

3.1. Origen de los datos

Los datos analizados en esta memoria hacen referencia a un curso online de programación en Android denominado "Aprende a programar tu primer..." . Este curso ha sido elaborado por profesores de la Escuela Politécnica Superior y se oferta a través de la plataforma edX de la que la UAM forma parte. Los cursos de edX se pueden ofertar en distintas modalidades. En concreto, el curso objeto de este estudio se ofrece de forma gratuita. Esta modalidad se denomina, tal como ya hemos visto, MOOC (Massive Open Online Courses). Lo que en general conlleva la matriculación de miles de alumnos...

Durante el análisis de los datos nos centraremos en el curso de programación Android. Sin embargo, se sigue un diseño genérico. Esto permite que los pasos detallados a continuación sean reproducirles en cualquier MOOC de la plataforma edX.

3.2. Información general

Comenzaremos hablando de los datos generales del curso, es decir, qué material se oferta y los distintos recursos disponibles en el curso.

3.2.1. Documentos

El curso se divide en 6 semanas, entre las que se reparte el temario dividido en 31 documentos. Cada semana cuenta con un documento donde se hace una breve introducción del nuevo tema. Esto hace en total, 37 documentos disponibles. La organización por temas y los nombres puede verse en el Anexo A:

3.2.2. Ejercicios

El curso se inicia con 7 preguntas sobre los conocimientos previos. Tras los primeros ejercicios hay 132 problemas que se van realizando durante el curso. Finalmente, hay 55 problemas que corresponden al examen final. Esto hace un total de 194 problemas. En el capítulo 4 analizaremos estos problemas y su relación con los vídeos.

Aunque hay 194 ejercicios, los datos muestran problemas numerados hasta el 196. Esto es debido a que hay dos problemas, el 139 y 140, que, si bien se crearon, no se mostraron a los usuarios, o se terminaron por eliminar. Y aunque en algún usuario puede constar como respondido, solo aportan ruido, por lo que se han eliminado de los datos.

Nota

Aunque la calificación obtenida por los estudiantes no depende solo de los ejercicios realizados durante el curso (7-132), sí que están muy relacionados. Estos solo suponen un 10 % de la nota final, pero permite relacionar temporalmente el avance del alumno, así como la relación con los vídeos. Por este motivo no nos centraremos en la nota global que obtiene los estudiantes del curso sino en la obtenida en los problemas.

Para simplificar los datos a aquellos problemas que estén formados por varias preguntas se les asignará una única respuesta. El problema se considerará acertado si se aciertan al menos la mitad de las preguntas que lo componen, y fallado en caso contrario. De esta manera podemos simplificar el análisis al tener todos los problemas como aciertos o fallos.

Vídeos

El curso cuenta con 30 vídeos distribuidos a lo largo del curso, además de un vídeo para la introducción de cada semana. Esto hace un total de 36 vídeos, pero para el análisis de los datos tan solo se usarán los vídeos donde se explica temario, es decir, se excluyen del análisis los vídeos de introducción, siendo 30 el total de los vídeos analizados.

3.3. Filtrado de los datos

Tras analizar previamente los datos podemos ver que, en muchas ocasiones, la información aportada carece de valor, o hay mucha diferencia entre las proporciones de un tipo particular de datos frente al resto. Solucionar esto supone tomar decisiones que permiten eliminar datos evitando perder la información relevante. La dificultad está en mantener generalidad, si los datos no contienen ejemplos representativos de las distintas situaciones, estos dejaran de representar la realidad. La existencia de datos atípicos no siempre es negativa. En el caso de algunos clasificadores, su aprendizaje es más robusto, y tolerable a ruido si durante el entrenamiento los datos no son perfectos. Para filtrar los datos vamos a seguir los siguientes enfoques:

3.3.1. Por usuario

Una forma sencilla seria no considerar todos aquellos usuarios que cuyos datos contengan ruido, o no se ajusten al esquema esperado. Pero esto supone eliminar demasiada información. En el curso actual contamos con poco más de siete mil alumnos, y son muchos menos los que terminan el curso.

Por ello vamos a evitar esta técnica en general. Sin embargo, en algunos casos carece de sentido mantener un usuario que solo aporta ruido. Este es el caso por ejemplo de un registro fallido que corresponde a un usuario sin nombre. Por ello estos datos no representan acciones de un alumno, sino un estado inconsistente en el log del curso. Por esta razón estos registros son eliminados, en el caso del curso actual esto supone un solo registro.

3.3.2. Por nota

Si el objetivo es eliminar a aquellos usuarios que no terminaron el curso, filtrar por nota parece la solución más simple, pero es difícil de determinar el umbral. Hay alumnos que aún terminando el curso no consiguen aprobar, y otros que, aun obteniendo buenas notas, abandonan antes de finalizar. Menos de mil alumnos realizaron el examen final. De modo que un umbral muy bajo no tendrá efectos en el filtrado, y muy alto hará que se pierda demasiada información. Dada la dificulta de ajustar un umbral óptimo, suponiendo que exista, filtrar por nota se descarta, y se utilizan otras alternativas.

3.3.3. Por número de eventos

Finalmente podemos eliminar los usuarios, analizando su comportamiento, que solo aportan ruido, entendiendo como ruido en el sentido por la dificultad de estimar sus patrones de comportamiento. Un usuario que no ve los vídeos realiza pocos ejercicios y termina abandonando tendrá muy pocos eventos generados. Por el contrario, un alumno constante, aunque no consigue los resultados obtenidos, solo viendo los vídeos, abriendo la documentación y realizando las preguntas obtiene un número bastante mayor de eventos. Estos son los usuarios que deseamos conservar. Generan información tanto de los vídeos visualizados como de las respuestas a los problemas.

Como en el caso previo, existe la dificultad de cómo determinar cuántos eventos son necesarios para pasar el filtro. Para analizar de forma visual el número de alumnos qué se mantiene según aumentamos el valor del filtro hemos generado una gráfica. La Figura 3.1 representa el número de usuario que se mantienen dependiendo del umbral que se establezca.

La Figura 3.1 ayuda a determinar el valor del filtro, este valor lo estableceremos en 150, es decir, no se considerarán aquellos usuarios que realicen menos de 150 eventos durante el curso. Este valor supone mantener tan solo el 25 % de los alumnos originales, pero dado que el 75 % eliminado registraban menos 150 eventos, el total de los eventos eliminados no es muy alto, pues representan menos de 35 % de los eventos totales. Es decir, tras realizar este filtrado conservamos el 65 % de la información original.

Para visualizar el impacto que suponen filtrar con este valor hemos generado las gráficas de las figuras 3.2, 3.3, 3.4 y 3.5. En todas estas figuras podemos ver la media de acierto, en porcentaje, de cada problema. Como ya hemos comentado anteriormente los problemas 139 y 140 se han eliminado de los datos. Dado que los 7 primeros problemas son un cuestionario sobre conocimientos previos y no representan un problema, sus valores son ignorados durante el análisis.

En las figuras 3.2 y 3.3 se utilizan solo los alumnos que hayan respondido explícitamente al problema para calcular la media. Comparando amabas figuras vemos que con independencia del número de estudiantes eliminados las proporciones no se han modificado apenas, es decir la proporción de aciertos y fallos se mantiene.

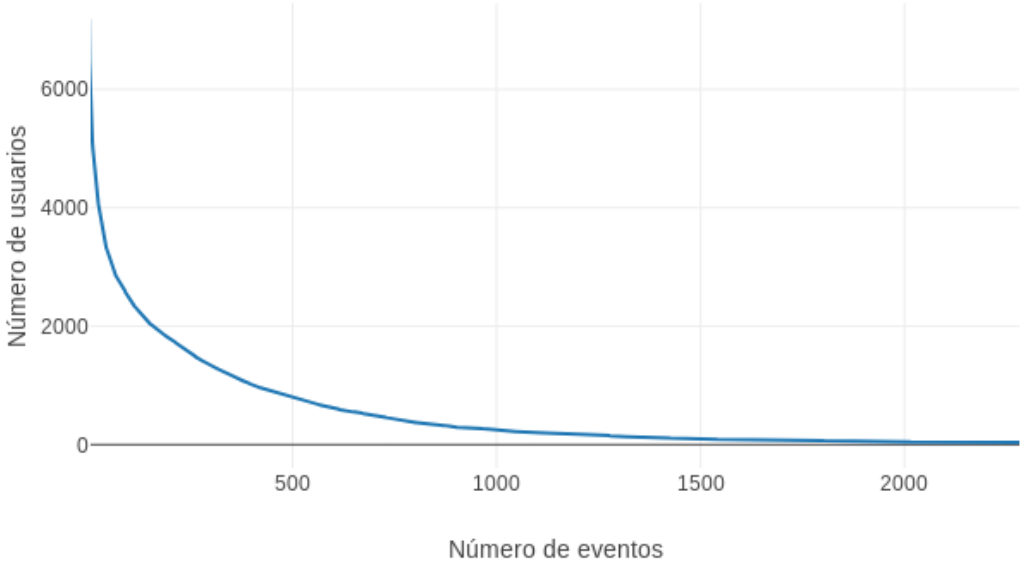

Figura 3.1: Número de usuarios que se mantiene según el valor del filtro

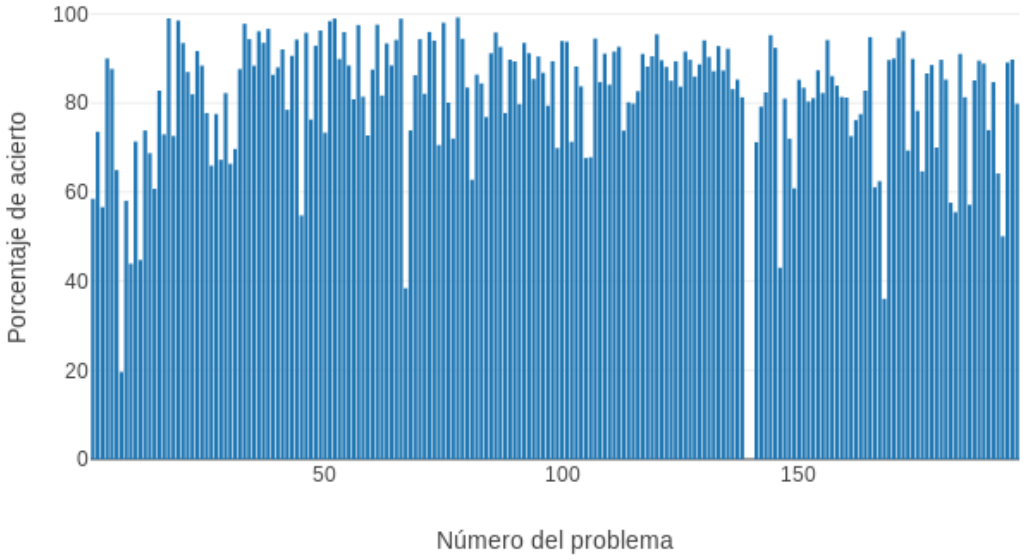

Figura 3.2: Media de los usuarios que respondieron sin filtro

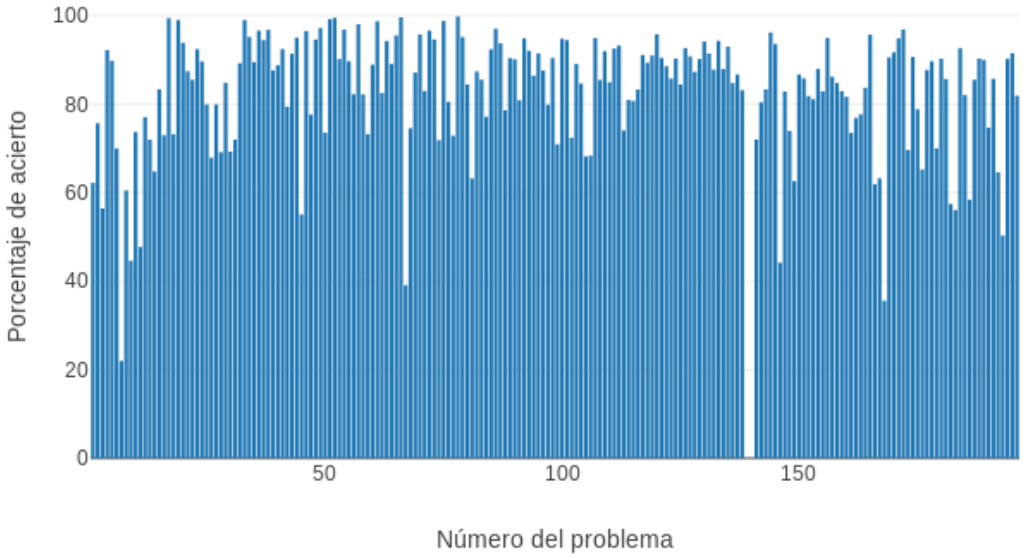

Figura 3.3: Media de los usuarios que respondieron con filtro

Para detectar que usuarios se han eliminado podemos observar las figuras 3.4 y 3.5. En estas gráficas no solo se tiene en consideración aquellos usuarios que responden explícitamente a los problemas sino que incluimos también a aquellos que no realizaron ninguna respuesta. Independientemente de si fue por abandono o no conocían la respuesta, asumimos que han respondido y fallado la pregunta.

Observando la Figura 3.4 vemos que sin filtrar los mejores problemas solo alcanzan el 30 % de acierto. Es decir, al incluir aquellos usuarios que no respondieron a la pregunta, la proporción de personas que fallaron el problema es mucho mayor que la proporción de aquellos que acertaron. Si comparamos la Figura 3.4 y la Figura 3.5 vemos que, tras realizar el filtrado de usuarios, las medias de acierto alcanzan valores del 80 %. Cifras que son más cercanas a los datos reales, es decir a los usuarios que respondieron explícitamente a la pregunta.

Esto implica que hemos eliminado gran parte de aquellos usuarios que iniciaron el curso, pero en las primeras semanas ya abandonaron, independientemente de su nota. Esto permite eliminar información difícil de analizar debido a su reducido número de interacciones con la plataforma online. Esto permitirá obtener mejore resultados que representen fielmente a los alumnos comprometidos con el curso.

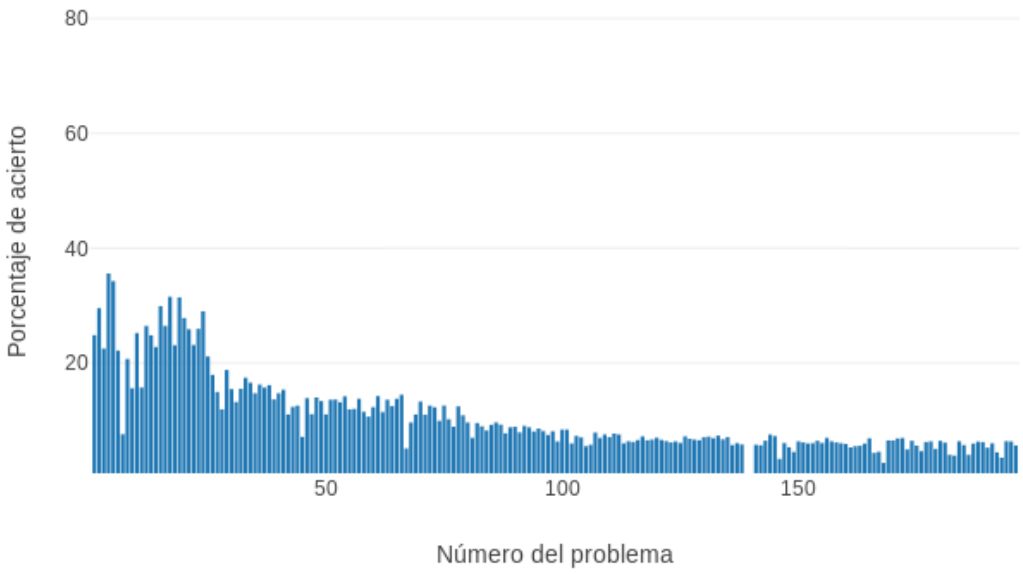

Figura 3.4: Media de los todos usuarios sin filtro

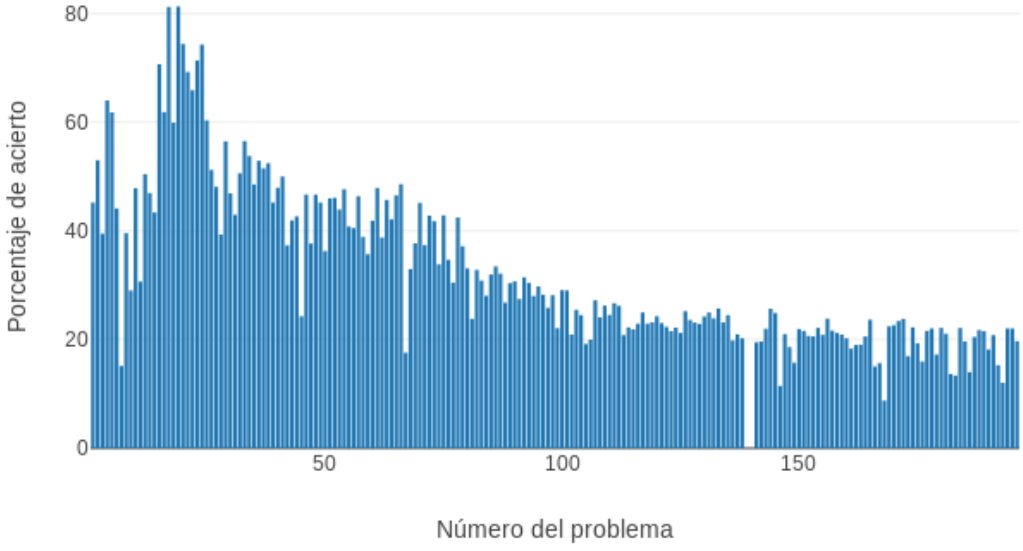

Figura 3.5: Media de los todos usuarios con filtro

3.4. Formato de los datos

En este trabajo partiremos de un fichero en texto plano que está compuesto de varias líneas en formato JSON. Cada línea contiene el número de usuario y una lista de sus eventos ordenados cronológicamente. El número de usuario es único, y todos los eventos de un usuario se recogen en una única línea. Durante el análisis no nos centraremos en los usuarios de forma individual, sino que se analizarán en conjunto. De esta manera podremos extraer características más generales. Este archivo proviene del Preprocesamiento realizado en "Predicción y análisis de interacciones de usuarios en plataformas de enseñanza online" [9], que transforma los datos generados en el log de la plataforma, en texto plano, al formato JSON que acabamos de mencionar.

Cada usuario realiza acciones durante el curso y estas son recogidas por los eventos. A continuación, detallaremos los distintos tipos de eventos, y explicaremos qué acciones los originan y qué implicaciones tienen en el análisis.

3.4.1. Tipos de eventos

1. **Eventos de vídeos:**

   Entre los distintos tipos de eventos que existen la gran mayoría tiene que ver con interacciones en la visualización de un vídeo. Todos los eventos de vídeo almacenan al menos la siguiente información: Id del vídeo que lo produce, fecha y hora del momento en que se genera el evento y el segundo del vídeo al que corresponde.

   a) **load_video:** El evento load_video es generado por el navegador al cargar un vídeo. Aunque presupone a un usuario que comienza a ver un vídeo, su generación es automática. Por tanto, depende del navegador y no del usuario. Por esta razón no se usará durante el estudio de los datos.

   b) **play_video:** El evento play_video es generado cuando el usuario comienza a ver un vídeo, o reanuda se reproducción tras una pausa o un stop. Este evento también puede generarse de forma automática, sin que el usuario haga clic explícitamente en el botón, cuando se produce un cambio del segundo de visualización (evento seek_video).

   Además, hay que tener en cuenta que, la aplicación permite descargar el vídeo completo para verlo posteriormente. Por lo que vamos a perder los eventos realizados durante la visualización offline del vídeo. Para evitar perder completamente esa información, la plataforma genera un evento de play_video en el inicio, y un stop_video en el último segundo. Esto supone, cada vez que un alumno se descarga el vídeo, en el log se genera una visualización sin pausas ni repeticiones del vídeo completo.

   c) **stop_video:** El evento stop_video es generado cuando se detiene la reproducción de un vídeo. Normalmente se genera cuando el usuario ha pulsado en stop. Aunque la mayoría de los casos esto significa que el vídeo ha terminado, no implica que el vídeo no pueda ser reanudado. Esto significa que no existe casi diferencia entre los eventos pause_video y stop_video. Por esta razón, cuando se haga referencia al evento pause_video se entenderá que también incluye el caso stop_video.

   d) **pause_video:** El evento pause_video es generado cuando el usuario pausa un vídeo. A diferencia del play_video y el stop_video, solo puede ser generado por el usuario. Nunca son generados como consecuencia de otra acción. Este tipo de evento solo puede ser generado cuando la reproducción del vídeo se realiza en la plataforma online.

e) **seek_video:** El evento seek_video es generado cuando un usuario se desplaza por la barra de visualización. Este evento además del segundo en el que se origina, también incluye el segundo al que se desplaza. Aunque aporte información relevante, este evento puede generar ruido. Si un usuario busca un segundo dentro de un vídeo, y este continua en reproducción, cada nuevo salto será un evento play_video en el nuevo segundo. Pero cuando se realiza el cambio, no siempre se genera un evento pause/stop desde el segundo donde se inició el movimiento, el segundo evento seek_video. Esto implica la aparición de dos eventos play_video seguidos. Estos eventos seguidos pueden generar picos en los segundo más visualizado por falsos play_video generados tras este evento.

f) **speed_change_video:** El evento speed_change_video se genera cuando la velocidad de reproducción es modificada. El evento también incluye la nueva velocidad indicada con un factor entre 0.5x y 2.5x. Generalmente es un evento poco frecuente, por lo que es difícil de utilizar para generalizar información.

2. **Eventos del foro**

   Estos eventos contienen la información de las acciones realizadas en los foros, ya sean foros generados por los profesores o alumnos. Dado que los esfuerzos en este TFG se centran en la relación entre problemas y vídeos, este tipo de eventos no se tendrán en cuenta en este trabajo.

   De forma similar a los eventos de vídeos, los eventos del foro almacenan al menos la siguiente información: Id del foro al que hace referencia y los detalles de la acción generada. Esta información extra puede ser el nombre del nuevo foro, la respuesta, o el texto buscado. Los distintos eventos del foro son:

   a) **edx.forum.comment.created:** El evento edx.forum.comment.created es generado cuando un nuevo foro es creado, este evento crea el id del foro, y contiene su nombre.

   b) **edx.forum.response.created:** El evento edx.forum.response.created es generado cada vez que se realiza una respuesta en un foro ya creado. Genera un evento de este tipo con el texto del comentario.

   c) **edx.forum.thread.created:** El evento edx.forum.thread.created es generado cuando se genera un nuevo hilo. Dentro de cada foro, se pueden albergar distintos hilos de discusión. Cada nuevo hilo genera en el usuario que lo ha creado un evento. En este se indica el foro al que pertenece y el nombre del nuevo hilo.

   d) **edx.forum.searched:** El evento edx.forum.searched es generado cuando se realiza una búsqueda en los foros. Esta queda registrada como evento con el texto buscado.

3. **problem_check:** El evento problem_check es generado cuando un estudiante responde a un problema. Generalmente los problemas constan de una sola pregunta, aunque algunos pueden estar compuestos por varias. En este evento se registra el id del problema, la fecha, el número de intentos, el número de preguntas que contiene y una lista con las respuestas. Cada respuesta incluye a su vez la puntuación obtenida y la máxima posible. Además, tiene un segundo id "natural"que se estableció manualmente en un fichero de configuración, durante la generación del archivo inicial. Este id representa el orden natural que un alumno debería seguir cuando responde las preguntas.

4. **openassessmentblock.self_assess:** El evento openassessmentblock.self_assess es generado cuando un alumno se autoevalúa una entrega. Durante el curso hay distintas prácticas que los alumnos deben autoevaluarse siguiendo una rúbrica e indicar qué nota han obtenido. Este evento recoge la nota que se asignan en cada parte de la práctica y la nota

máxima posible, así como el id de la misma. Dado que la evaluación no es revisada, un alumno puede asignarse una nota que no se corresponda con la rúbrica. Como esta nota no es un dato objetivo no se usará en los análisis.

5. **textbook.pdf.chapter.navigated:** Además de los vídeos, foros y problemas, los alumnos también disponen de documentación en formato PDF. Cada vez que un alumno se descarga o abre uno de estos documentos se generará un evento con la fecha y el nombre del documento.

6. **error_json:** Estos eventos son ignorados. Se trata de casos muy raros y suponen que, durante la lectura del log, algún carácter o registro incoherente, generó un error. Este error impidió obtener toda la información y de esta forma queda registrado el error.

3.5. Generación de secuencias de visualización

Analizados los datos originales y los distintos tipos de eventos, vemos la necesidad de crear un nuevo tipo de datos que facilite el estudio y análisis de los vídeos. Estos datos son las "Secuencias de visualización". Estos datos representan la reproducción de un trozo del vídeo y son generados a partir de los eventos play_video y stop_video. Incluyen la información del segundo de inicio y final. Así como la fecha y hora de los eventos que generaron el inicio y final de la secuencia.

Adicionalmente durante la generación de estas secuencias de visualización se almacenarán otros datos extra que servirán, más adelante, para optimizar la obtención de resultados. Estos son, por ejemplo, la fecha que se respondió por última vez un ejercicio o la duración de los vídeos. Para generar las secuencias de visualización se tendrán en cuenta distintas consideraciones:

Por cada usuario se separarán los eventos de play, pause y stop del resto. De nuevo, estos eventos se volverán a separar por id del vídeo, separando los distintos vídeos del curso. Esta última separación se realiza para evitar errores de concurrencia. Por ejemplo, al tener dos pestañas abiertas con dos vídeos se pueden distorsionar los datos si se registra un evento pause_video del vídeo B, entre el evento play_video y pause_video del vídeo A.

Los eventos se analizan cronológicamente, teniendo solo información del evento actual y el inmediatamente anterior. Como indicamos anteriormente, los eventos stop_video y pause_video se tratarán del mismo modo.

Un evento reproducción comienza con un play_video y termina con un pause_video. Para entender el orden esperado de los eventos hemos generado la Figura 3.6. El vídeo comienza con un evento play en el segundo 0 y, a continuación, se produce un evento de pause que genera la primera secuencia de reproducción, desde el segundo 0 al 2. Tras el pause, en el mismo segundo, se genera un evento play, y crea una nueva secuencia que termina con el pause del segundo 5. Entonces el usuario vuelve al segundo 2 (evento seek_video) y reproduce de nuevo un trozo de vídeo, desde el segundo 2 al 9. Tras esta última parada, visualiza el vídeo hasta el final y termina con un stop. De manera que este vídeo genera 4 secuencias de visualización: 0-2, 2-5, 2-9 y 9-11.

La Figura 3.6 sería el caso perfecto. El vídeo comienza con un play y termina en un stop, tras un play hay siempre un pause/stop, se pausa el vídeo antes de cambiar de segundo y se vuelve a reproducir tras el cambio, y los plays tras los pauses se generan en el mismo segundo que el último play.

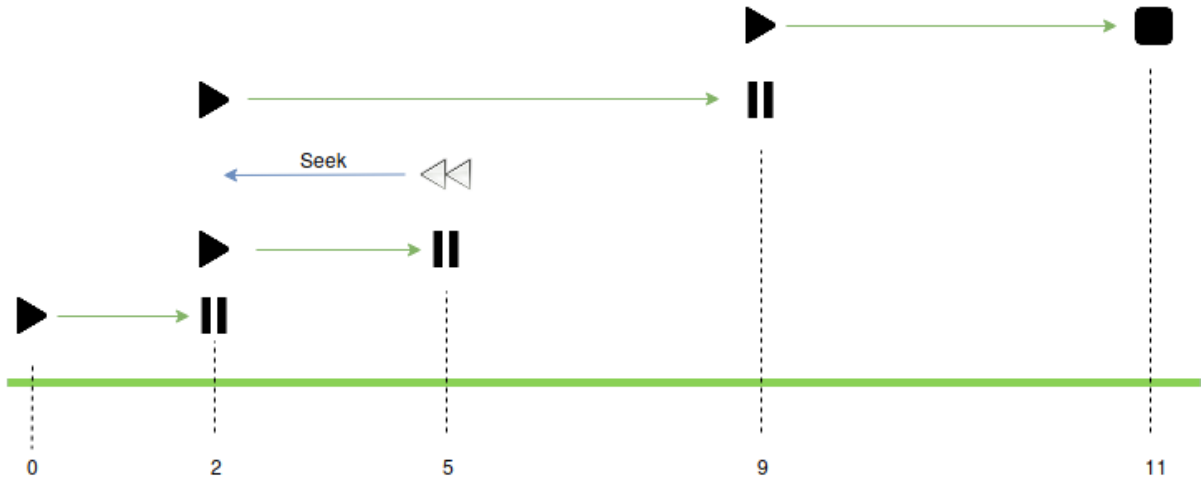

Figura 3.6: Caso esperado en la reproducción de un vídeo

Sin embargo, hay pocos casos en los que los datos son tan perfectos, en la Figura 3.7 hemos agrupado los distintos problemas que nos podemos encontrar. El vídeo comienza normal, pero tras el primer pause en el segundo 2, no hay un evento play, sino que el siguiente evento es otro pause en el segundo 5. No podemos asumir que en el segundo 2 hubo un evento play, por lo que no se genera una secuencia de reproducción. En el segundo 5 tras un play, encontramos otro en el segundo 9. En este caso si hemos considerado que entre ambos eventos si ha habido visualización del vídeo, debido en que algunos navegadores pueden generar automáticamente eventos de play durante la reproducción para tareas como son la sincronización del vídeo. Pero solo consideramos que hay visualización si el tiempo en el vídeo avanza y no retrocede, por ejemplo tras el play del segundo 9 hay otro en el segundo 5, carece de sentido que el vídeo se reproduzca hacia atrás, así que podemos asumir que hay un evento seek no registrado. Pero es imposible saber desde donde se originó por lo que esa información es imposible recuperarla, pero al menos mantenemos los datos que se generen con el segundo play. Finalmente hay un evento stop, que podría ser un pause también, con el que se genera una nueva secuencia de reproducción desde el segundo 5 al 11. En total se han generado 3 secuencias de reproducción: 0-2, 5-9 y 5-11.

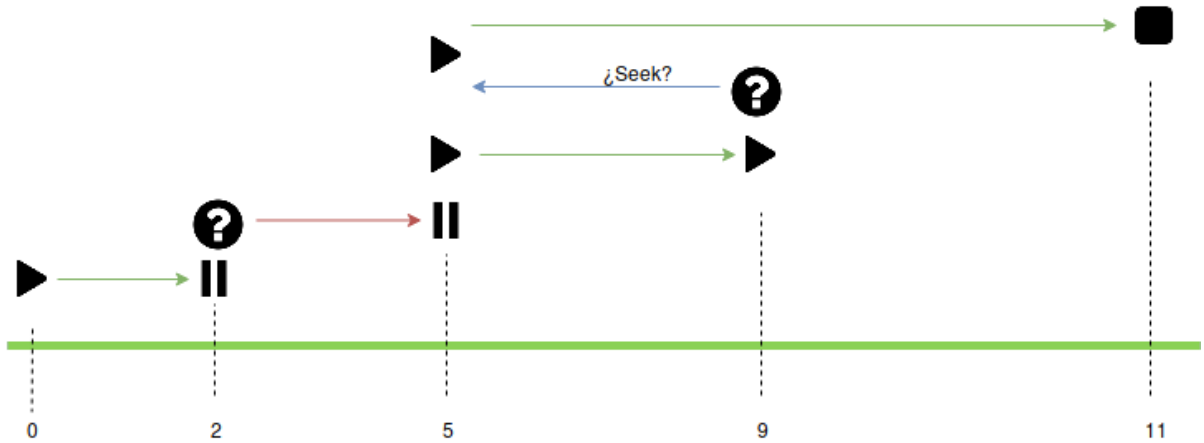

Figura 3.7: Caso real con errores en la reproducción de un vídeo

En definitiva, cuando dos eventos pause_video se dan seguidos no se genera una secuencia de reproducción. El primer pause_video finaliza la secuencia anterior, si la hubiese y no se comienza una nueva hasta que aparece un nuevo evento play_video. Cuando una secuencia de reproducción ha empezado, y tras el primer evento play y aparece otro evento play_video, se asume que antes del segundo evento play_video hubo un evento pause_video. Esto implica que finaliza la secuencia de reproducción anterior y comienza una nueva.

Para evitar las suposiciones, en el caso de dos plays seguidos, se ha intentado comparar el tiempo transcurrido entre la fecha y hora que se generaron los eventos con el tiempo transcurrido entre el segundo inicial y final de reproducción del vídeo. Sin embargo, más del 30 % de los datos superan una diferencia mayor a 5 segundos entre ambos tiempos. Esto se debe a que las fechas de los eventos dependen de factores externos como tiempo de respuesta, acceso a base de datos y otras circunstancias que ni son deterministas, ni dependen del usuario. Por esta razón no se ha podido utilizar esta comparación para determinar cuándo dos plays seguidos son válidos, de modo que se ha considerado que dado dos plays seguidos con tiempo consecutivo se considera siempre como válido.

El ruido es un problema en todos los preprocesamiento de datos, tras la transformación en el log [9], se elimina el ruido ajeno a los datos, causa de registros incoherentes del log de la plataforma edX. Pero no se elimina los problemas que los propios datos almacenan, que son los errores que hemos mencionado.

Tras realizar este procesamiento, entre todos los usuarios y vídeos obtenemos en total 357.123 secuencias de reproducción. Aunque es un número aparentemente grande, hay que contar con miles de alumnos que realizan miles de eventos, es decir, se supera la cifra del millón de eventos en total. Estas secuencias concentrar toda la información y esta ordenada por usuario y vídeos, de modo que analizar la visualización de un vídeo resulta muy rápido y no es necesario procesar todos los eventos del curso cada vez que se necesite esta información.

Para ver que los datos no son perfectos. A continuación, mostramos los problemas encontrados, en qué cantidades, y cómo se han tratado:

1. Es posible encontrar registros donde tiempo del vídeo era “None”, es decir, no tenía tiempo. En nuestro curso, este error aparecía una sola vez y este tipo de eventos son ignorados.

2. Del total las secuencias de reproducción generadas, 119.345 (30 %) son generados mediante 2 play_video seguidos. Este es el caso en el que se asumía que antes del segundo play hay un pause y generaba una nueva secuencia de reproducción.

3. Se han encontrado un total de 54.813 eventos play que iban precedidos por un evento play cuyo tiempo era posterior al segundo play. Como hemos indicado anteriormente el primer play es ignorado, por desconocer cuándo terminaría la secuencia de reproducción.

4. Se han encontrado 50.211 eventos pause_video sin un evento play_video previo. Es decir, dos pause_video seguidos. El primer pause termina la secuencia de reproducción empezada en el play anterior, si lo hubiera, y el segundo pause es ignorado.

4

Análisis de los datos

En este capítulo analizaremos los datos generados. Primero introduciremos las gráficas que genera cada vídeo con sus visualizaciones. Continuaremos analizando qué segundos son más importantes en los vídeos. Y, finalmente, veremos cómo relacionamos los vídeos y problemas. El curso cuenta con más de 30 vídeos y más de 190 problemas. Dado que son demasiadas gráficas nos centraremos explicar aquellas que por sus características merecen ser destacadas.

4.1. Visualización de los vídeos

En esta sección vamos a ver las gráficas con los datos de las reproducciones de los vídeos. En el anexo B se pueden encontrar estas mismas gráficas con distintos filtros según el número de eventos por usuario. Pero en este capítulo nos centraremos tan solo en las gráficas de los datos que han sido filtrados los usuarios con menos de 150 eventos. En el anexo B se puede observar que, cuanto más alto es el umbral del filtro, más similares son los datos de los distintos vídeos. Dado que es el único detalle de importancia, todas estas gráficas no se muestran en este capítulo.

4.1.1. Número de visualizaciones totales

El número de visualizaciones totales nos permite saber cuántas veces se ha visualizado un segundo concreto de un vídeo. De este modo podremos detectar qué segundos se visualizan más. También permite analizar el comportamiento de los usuarios según avanza el vídeo. Por ejemplo, si el número de visualizaciones de la parte final del vídeo es el mismo que el de la parte inicia, si ha crecido o si ha descendido y cuánto.

Todas las gráficas mostradas se generan en fichero HTML, que más adelante veremos en qué ventajas supone frente a una imagen estática. En la Figura 4.1 se representan el número de visualizaciones totales de cada instante del vídeo para los 30 vídeos del curso. El instante de tiempo de los vídeos esta normalizado de forma que 1 comprende el final de todos los vídeos y 0 el inicio. Esto se hace para evitar que dependiendo de la duración del vídeo la traza de un vídeo termine antes o después. Este eje por tanto representa la fracción del tiempo del vídeo. Como se puede ver en la Figura 4.1 si se muestran los datos de todos los vídeos, resulta difícil interpretarlos. Por esta razón en la Figura 4.2 se muestran algunos de ellos.

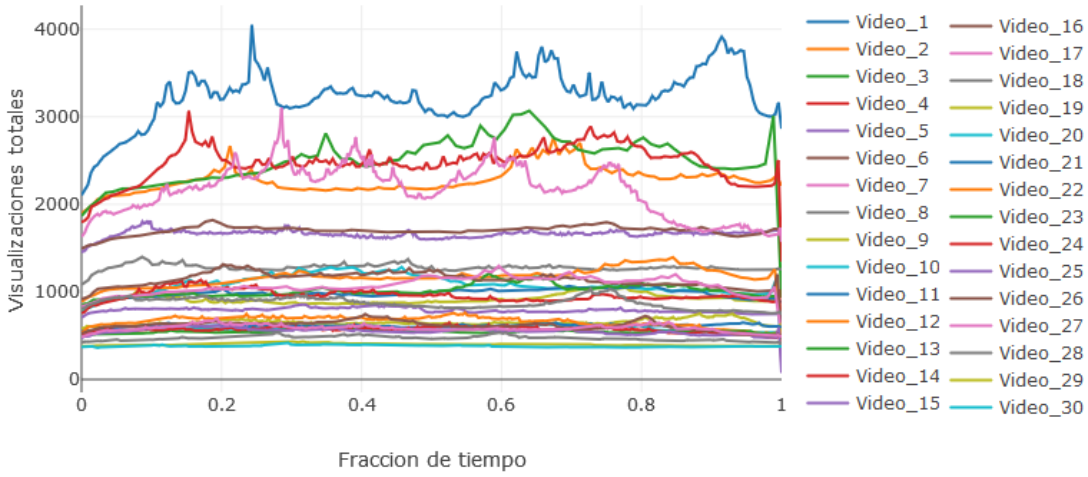

Figura 4.1: Número de visualizaciones totales

En la Figura 4.2, se muestran los vídeos 1, 2, 4, 6, 17, 26 y 30, por ser representativos del curso. Como se pude observar los primeros vídeos tienen muchas más visualizaciones. Esto se debe a que a pesar de haber filtrado usuarios, aún hay muchos que abandonan o dejan de ver los vídeos según avanza el curso. Es importante resaltar que no se observa una disminución en el número de visualizaciones de los vídeos. También puede observarse que hay secciones de los vídeos más vistas que otras.

Normalización

En la Figura 4.1 y la Figura 4.2 se muestra el número de visualizaciones totales. Esto implica que no se aprecie con detalle las diferencias en los vídeos con menos reproducciones.

Para apreciar mejor estas diferencias, se puede normalizar el valor del eje y, visualizaciones totales. Aunque normalmente se utilice la media y desviación, en este caso es mejor usar el máximo. Dado todos los valores para cada instante de tiempo de un vídeo, se obtiene el valor máximo y se dividen todos los valores por él, de esta forma conseguimos que todos estén entre 0 y 1.

Tras esta normalización, en la Figura 4.3 se puede apreciar mejor qué instantes, o segundos, son los más reproducidos. Al igual que en la Figura 4.2 se han mostrado solo los vídeos más representativos.

Figuras como estas permiten ver qué segundos son aquellos que los usuarios necesitan ver más veces. De este modo podemos ver qué partes del vídeo son problemáticas. Ya sea porque su explicación no es clara, porque contiene varios conceptos clave o porque detalla cómo realizar cierta actividad.

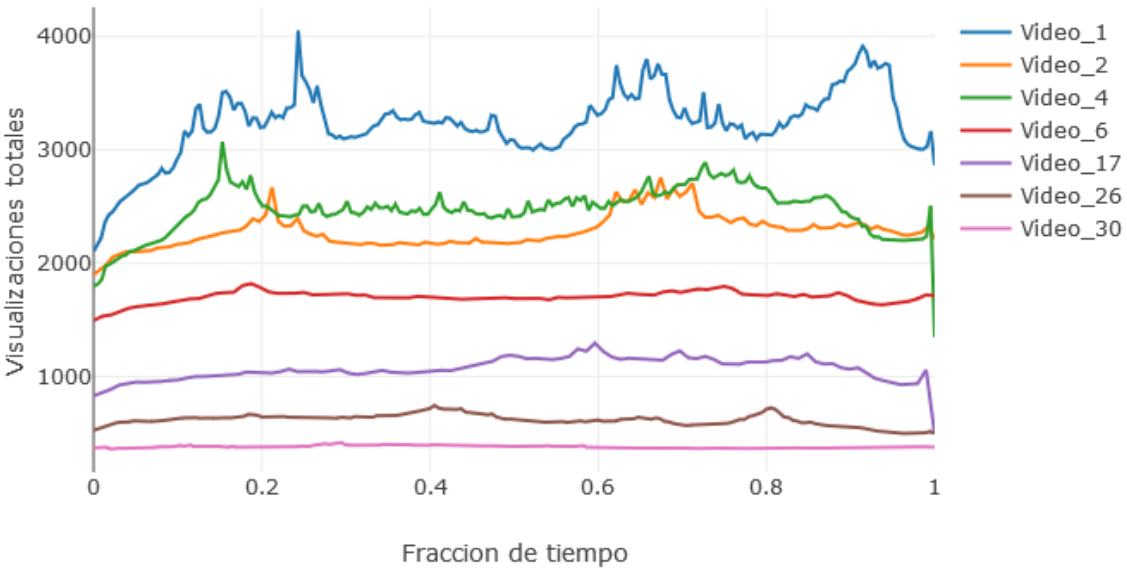

Figura 4.2: Número de visualizaciones totales (Vídeos representativos)

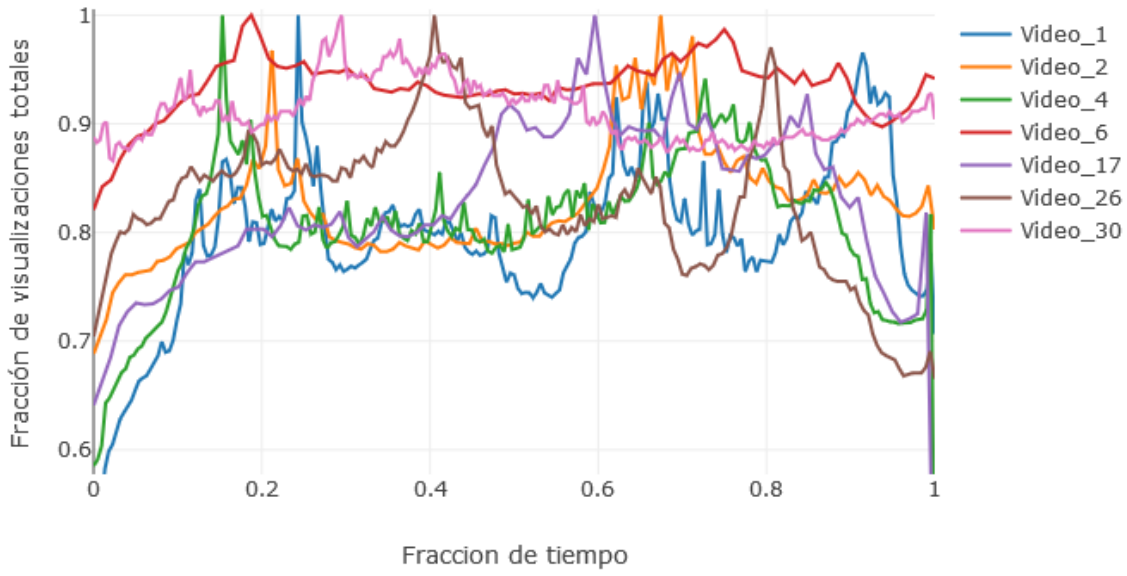

Figura 4.3: Fracción de visualizaciones totales (Vídeos representativos)

Por ejemplo para el vídeo 2 se puede observar que los últimos segundos son mucho más visualizados. Dado que los segundos han sido normalizados entre 0 y 1 se desconoce el segundo exacto al que hacen referencia los valores del eje x. Pero conociendo la duración del vídeo puede obtenerse este. En este caso, el vídeo 2 dura 138 segundos, es decir si los porcentajes más visualizados están entre el 60 % y 80 % del vídeo, significa que los segundos más visualizados están entre el segundo 82 y 110. Si se desea evitar esta conversión, durante la generación de las gráficas se puede indicar que no se normalice el eje x, y que este corresponda directamente a los segundos. El resultado al no aplicar esta normalización puede verse en la Figura 4.4.

Al no normalizar los vídeos, aquellos más largos reducen el espacio en la gráfica de los más cortos. Por ello, que, para analizar rápidamente todos los vídeos en una sola gráfica, es preferible normalizar ambas dimensiones, y si se desea un análisis más profundo, generar las gráficas con los valores absolutos.

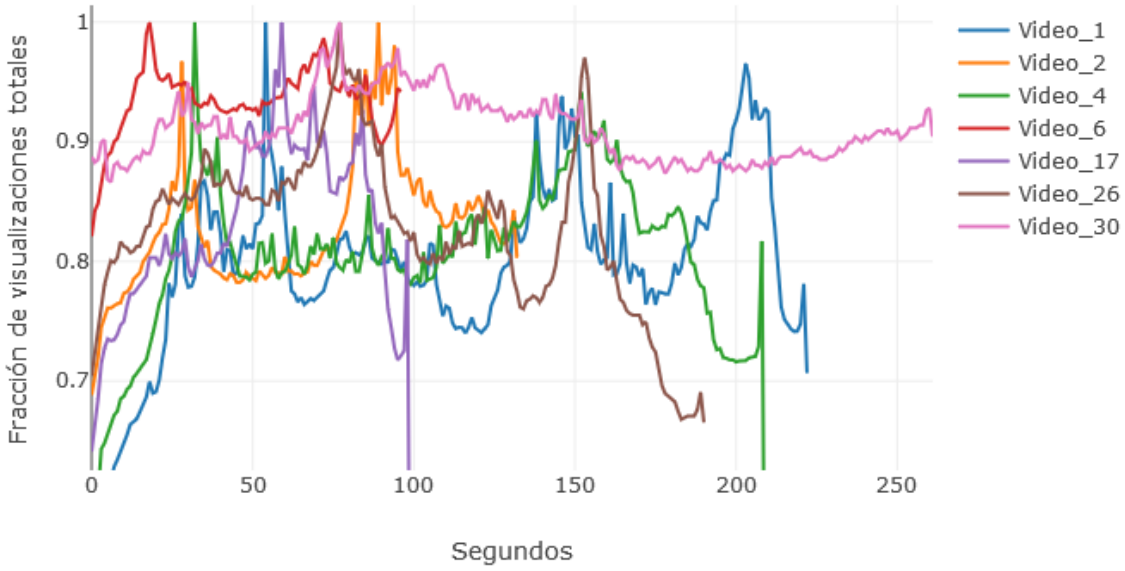

Figura 4.4: Fracción de visualizaciones totales en función de los segundos (Vídeos representativos)

4.1.2. Número de usuarios con visualizaciones

En el apartado anterior nos hemos centrado en las visualizaciones totales por vídeo. Sin embargo, un vídeo puede ser reproducido por un mismo usuario varias veces mientras que otro usuario no lo vea. Para poder analizar las visualizaciones por alumnos generamos nuevas gráficas. Esta vez el eje y no contiene el número de visualizaciones totales sino el número de usuarios que han visto cada instante de tiempo del vídeo cómodamente. Y en el eje x se mantiene la fracción del tiempo del vídeo.

El resultado se muestra en la Figura 4.5, donde descubrimos que la visualización de los vídeos no disminuye según avanza el tiempo. Sino que aquellos usuarios que comienzan un vídeo suelen terminarlo. También podemos observar que muchos usuarios se saltan los primeros y últimos

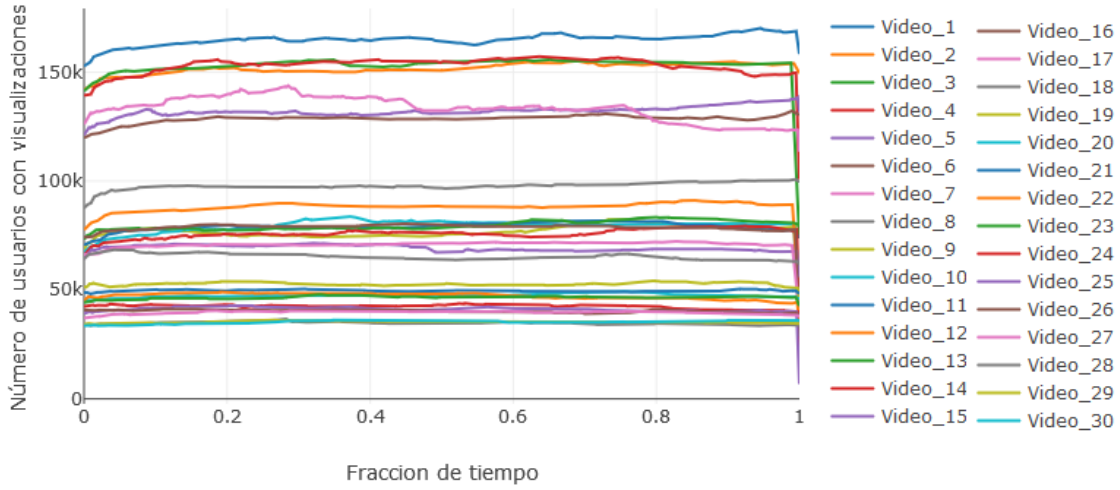

Figura 4.5: Número de usuarios con al menos una visualización por instante de tiempo

segundos de cada vídeo. Gracias a esta gráfica podemos determinar si en algún vídeo, existe una zona donde el número de alumnos deje de visualizar el vídeo, en este curso, no se aprecia ese comportamiento en ninguno de los vídeos.

En este caso al tener los valores absolutos, se puede observar tres "zona", o agrupamientos de vídeos por número de usuarios con visualizaciones. Estas zonas tienen correspondencia temporal con las semanas del curso. La primera zona corresponde a los vídeos de la semana 1 (vídeos 1 a 7). Dentro de esta zona podemos distinguir como los tres primeros vídeos son los más visualizados. Tras la primera semana el abandono es notable. Casi la mitad de los usuarios que vieron el primer vídeo no ven los vídeos de la segunda semana (segunda zona, vídeos 8 a 14). De nuevo, tras la segunda semana, se observa un notable descenso en la tercera semana y posteriores, que constituyen la tercera zona (vídeos del 15 en adelante). Como se puede apreciar en la gráfica, no es hasta la tercera semana que el número de usuarios se mantiene constante.

Tan solo unos 700 alumnos, el 10 % del total, visualizan todos los vídeos. Esta cifra coincide aproximadamente con el número de usuarios que realizan el examen final. De esta manera este gráfico puede ayudarnos ya en la tercera semana a estimar el número de alumnos que terminarán el curso.

Normalización

Si deseamos evitar que los vídeos se diferencien por número de usuarios totales podemos aplicar una normalización similar a las gráficas de visualizaciones totales. En este caso se obtendría el valor del instante de tiempo en el que hay más usuarios viendo el vídeo, y se dividen todos los valores por el máximo obtenido. Sin embargo, como vemos en la Figura 4.6, el efecto no es el esperado, ya que todos los vídeos son muy parecidos. Además, al juntarse todas las trazas, su visualización es más difícil, y los pequeños detalles se aprecian demasiado. Si un usuario ha visto solo algunos segundos del vídeo, observaremos picos, pues en algunos vídeos tratamos con tan solo 400 usuarios.

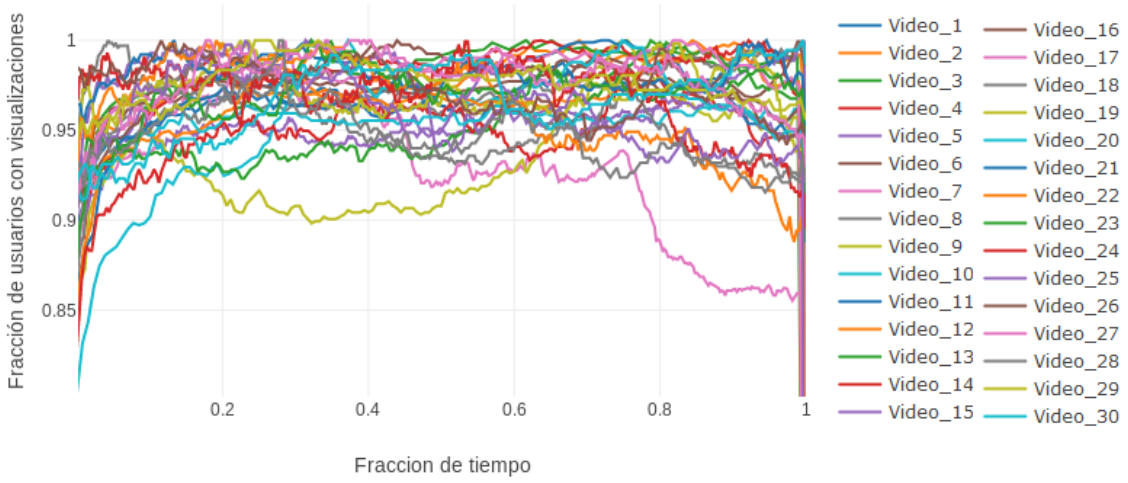

Figura 4.6: Fracción de usuarios con al menos una visualización por instante de tiempo con respecto al máximo de estudiantes que han visualizado a la vez el vídeo en un mismo instante

Al normalizar a 1 el valor máximo todo el grafico de la Figura 4.6 centra sus valores entre 0,8 y 1. Es por esto que al centrar el grafico en una zona tan pequeña los detalles resaltan demasiado, y tomando más importancia datos individuales que la información del general.

4.1.3. Análisis de zonas críticas

Las gráficas de visualizaciones totales permiten ver qué segundos son más reproducidos. Las gráficas con el número de usuarios ayudan a identificar problemas con el abandono. Si queremos ubicar los puntos más problemáticos de un vídeo, no basta con saber qué partes son más visualizadas. Por ejemplo tras una parte complicada, es común pausar el vídeo para analizar lo que se ha explicado, o desplazarse en la barra de visualización repetidamente buscando una zona en concreto. Para analizar el número de interacciones que hace el alumno con el vídeo vamos a realizar un nuevo tipo de gráfica.

De nuevo todos los vídeos en una única gráfica imposibilita el análisis. Por ello hemos escogido aquellos que hemos considerado más importantes, ya sea por su trazado o por sus características. Los vídeos seleccionados para el análisis son 2, 3, 6, 21, 27 y 30.

Estos vídeos se analizan en la Figura 4.7. En el eje x se representa el avance temporal del vídeo, y en el eje y, el porcentaje de eventos realizados hasta ese momento. Ya en el segundo 0 hay numerosos eventos de play, pues estos se generan automáticamente al comenzar un vídeo.

A continuación, mostramos algunas de las zonas que podemos analizar en la Figura 4.7. Si un vídeo se reproduce sin pausas este registrará la mitad de sus eventos en el instante 0 (plays) y la otra mitad en el instante final (stops). Por ejemplo el vídeo 6 apenas registra nuevas acciones durante su finalización, esto implica que casi todas las interacciones con el vídeo suceden en la primera mitad. Por el contrario, el vídeo 3 no tiene muchos eventos en su inicio, pero estos crecen durante el final. Lo ideal es que el temario este bien repartido, es decir su crecimiento sea lineal, como por ejemplo en el vídeo 21. También es deseable evitar que una pequeña zona concentre muchos eventos, como sucede en el vídeo 2, que acumula una gran cantidad de eventos hacia la

mitad del vídeo. Esto implica que esa zona resulta confusa, o condensa demasiada información que debería estar más repartida.

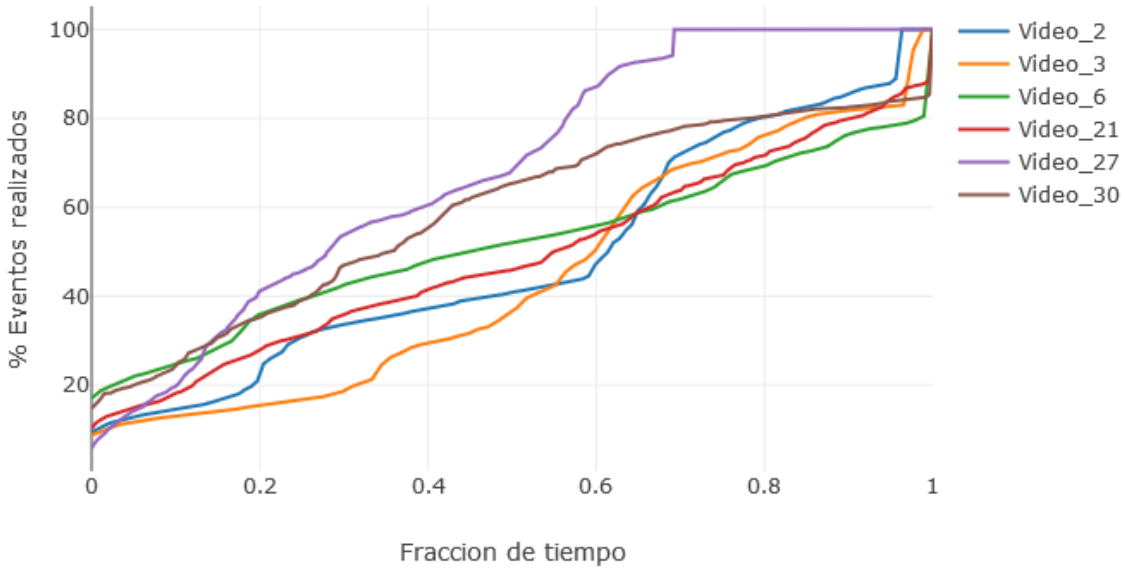

Figura 4.7: Fracción de eventos acumulados en función del instante (Vídeos representativos - Zonas críticas)

Analizando cada vídeo individualmente podemos determinar cuáles son más problemáticos, obteniendo qué puntos se deben analizar, e identificar el problema. Dado que son 30 vídeos no nos hemos centrado en determinar cuáles son estas zonas de cada vídeo, sino en cómo descubrirlas y poder interpretarlas.

4.2. Relación vídeos y problemas

El objetivo de los vídeos es trasmitir conocimiento y los problemas comprueban si los conocimientos adquiridos son correctos. Por tanto, el siguiente paso consiste en utilizar los problemas para analizar si los vídeos transmiten correctamente los conocimientos explicados.

4.2.1. Metodología

Para obtener esta correlación utilizaremos algoritmos de predicción. Es necesario conocer la relación de cada vídeo con todos los problemas. Lo que supone, teniendo alrededor de 30 vídeos y casi 200 problemas, realizar 6000 ejecuciones. Si bien es la parte más costosa del proyecto, su tiempo de ejecución en un ordenador de usuario no supera la media hora.

Para obtener esta relación intentaremos predecir el resultado de un problema conociendo el patrón de visualización del vídeo relacionado. La predicción consiste en acierto o fallo de cada problema por usuario, y la entrada son las visualizaciones de cada segundo del vídeo. Para evitar el sobre ajuste y reducir el número de argumentos en la entrada, se agrupan en periodos

de 20 segundos las visualizaciones. Una primera agrupación empezando en el inicio y con saltos de 20 segundos. Y una segunda agrupación comenzando en el segundo 10 y con saltos de 20 segundos, eligiendo 10 por ser la mitad del intervalo de los saltos.

4.2.2. Algoritmos

Una vez determinada la entrada y salida, es necesario determinar qué algoritmo utilizar en la predicción, además de sus parámetros. Durante la ejecución se utiliza validación cruzada para evitar fluctuaciones estadísticas. Es decir, obtener la media de las predicciones, y aumentar la robustez de los resultados.

Algoritmos de predicción

El objetivo no es realmente obtener el resultado de los problemas a partir de los vídeos, sino ver cómo de relacionados están y como medida de correlación se usará la tasa de acierto del algoritmo de predicción. Para determinar cuál es el mejor, se han probado distintos algoritmos de predicción y variado los parámetros propios de cada algoritmo. Entre los algoritmos probados se encuentran:

1. Vecinos próximos (KNN)
2. Maquinas de soporte vectorial (SVN)
3. Gaussian processes (GP)
4. Árboles de decisión
5. Random Forest
6. Redes neuronales multicapa (MLP)
7. Naive Bayes
8. Quadratic discriminant analysis (QDA)

Algunos algoritmos fueron descartados por su alto coste computacional. Tras varias modificaciones en los modelos usados y atributos de configuración, el algoritmo que mejores resultados obtenía ha sido el de árboles de decisión. Agrupar los segundos en periodos de 20, permite obtener buenos resultados con tan solo una profundidad máxima de 5 niveles. Esto ayuda a reducir tanto el coste de entrenamiento como coste para obtener los resultados.

Algoritmos de selección

Una vez obtenidos los resultados, es necesario decidir si es posible extraer información útil. Por ejemplo si un problema tiene el 90 % de acierto, no es suficiente que el clasificador acierte en el 90 % de los casos. Además si independientemente del vídeo los resultados son similares, quiere decir que la predicción se basa en el porcentaje de las clases y no en los vídeos. Igualmente se han revisado las matrices de confusión en algunos resultados seleccionados aleatoriamente, asegurando que los fallos en ambas clases sean proporcionales a los datos.

Estos problemas se pueden ver más adelante en las gráficas, con lo que denominaremos "cresta" continuas en los datos de un problema. Analizando la desviación media de los resultados

de los distintos vídeos en un mismo problema puede detectarse. La desviación será muy baja cuando los resultados sean similares, es decir la tasa de acierto obtenga valores próximos entre sí en todos los vídeos, por lo que el patrón de visualizaciones no es relevante en ninguno de los vídeos. Analizando el los resultados de las desviaciones obtenidas hemos considerado adecuado fijar este valor en 10^{-10}. Este valor permite detectar errores en los datos como el problema 139 y 140, es decir, detecta cuando no existe ninguna relación entre los vídeos y el problema. Pero no detecta los datos en los que se puede apreciar, visualmente, que los resultados son muy similares en los distintos vídeos.

Una vez obtenidos los resultados de ejecutar el algoritmo de predicción en necesario determinar cómo asignar las relaciones. Todos los problemas deben explicarse en algún vídeo, pero la explicación podría repetirse en varios vídeos. De modo que se analiza cada problema individualmente y se obtiene qué vídeo maximiza la tasa de acierto en la predicción, con lo que asociamos el vídeo al problema.

Si queremos evitar asignar un único vídeo a cada problema, pero tampoco queremos asignar siempre "X" vídeos, podemos relacionar todos los problemas que superen un umbral fijado. Pero dependiendo del problema la media y el máximo de los resultados puede ser muy diferente. De modo que el umbral debe ser diferente en cada problema, este umbral será cuantos errores más que el mejor resultado consideramos como válidos, es decir, si el mejor resultado obtuvo 97 aciertos de 100 datos, y nuestro umbral es 2, relacionaríamos todos los vídeos que al menos hayan acertado 95 veces en el test.

Dado que tras la ejecución solo conocemos los porcentajes de acierto, para calcular que porcentaje representa el umbral utilizamos la formula 4.1. Al valor del resultado máximo obtenido en la predicación de un problema se le resta el porcentaje correspondiente al número de fallos extra que toleramos. En la formula 4.1 el denominador es el tamaño del test en el algoritmo de predicción. Mientras que el numerador, "A", es el valor de nuestro umbral.

$$Porcentaje(A) = Max - A/(\frac{\text{Tamaño de los Datos}}{Particiones}) \tag{4.1}$$

Pongamos un ejemplo usando la formula. Si tenemos que la tasa máxima de acierto es 51 resultados en un conjunto de 100 datos, es una tasa de acierto del 51 %. Si queremos aceptar todos los resultados con máximo 2 fallos más, necesitamos saber qué porcentaje representan esos 2 datos. Es decir, 2 datos de 100, el 2 %. 2 es valor que asignamos al umbral y 100 el tamaño del test, que es el número de datos totales dividido por las particiones de la validación cruzada. Restando al máximo ese porcentaje permitimos que todos aquellos valores por encima de este porcentaje tengan como máximo 2 fallos más que el mejor resultado.

Estos datos se utilizarán para la generación de nuevas gráficas que permitirán comparar el comportamiento de los usuarios que acierten y los que fallen. En estas gráficas se representarán solo aquellas predicciones que tengan una tasa de acierto mayor al resultado de la función según el umbral definido.

4.2.3. Parámetros del análisis

El objetivo no es obtener unos datos, analizarlos y exponer los resultados, sino la creación de una herramienta con los algoritmos necesarios para poder analizar e investigar sobre cualquier tipo de datos. Para ello todo el diseño busca la simplicidad y parametrización de los algoritmos. Los parámetros disponibles son:

1. Número de eventos mínimo para filtrar usuarios.

2. Fichero que se desea analizar.

3. El valor mínimo de la desviación estándar para considerar un problema como no válido. Si la desviación de algún problema tiene un valor menor a este umbral se indicará durante la ejecución del análisis de los datos.

4. Lista de problemas a excluir. Si se desea eliminar del análisis, podemos evitar que se visualicen los resultados de aquellos problemas indicados en esta lista. Este es el caso de los ejercicios 139 y 140.

5. Incluir los usuarios que no han hecho un problema como respuesta fallada. Se considera a los usuarios que no responden a un problema, de modo equivalente a responderlo al final del curso y fallar su respuesta. No incluir estos datos hace que en algunos problemas del final de curso la cantidad de alumnos sea demasiado pequeña para obtener resultados analizables.

6. Umbral de relación. Como hemos indicado antes, este valor permite relacionar más o menos vídeos a un mismo problema. Cuanto más alto sea este valor, más vídeos alcanzarán la tasa de acierto necesaria para considerar que hay relación entre el vídeo y el problema. Si su valor es 0 solo el vídeos que tengan el valor máximo son relacionados, pudiendo ser varios si tienen la misma tasa de acierto.

7. Fecha de realización del problema. Los datos de visualización en un vídeo usados en los algoritmos de predicción solo incluyen desde el inicio del curso hasta la fecha de realización del ejercicio que se desea relacionar. Sin embargo, si incluimos los estudiantes que no respondió un problema, no se puede determinar la fecha de respuesta. La fecha que se establece es el final de curso. Es decir, todas las visualizaciones de un vídeo durante el curso se incluyen. Para eliminar esta diferencia, se puede evitar limitar los datos de aquellos que realizaron el problema, a la fecha de realización, de modo que se incluyen todas las reproducciones del curso también a aquellos usuarios que respondieron el problema.

8. Intervalo de agrupación. Para reducir el número de entradas en los algoritmos de perdición, la entrada se agrupa en intervalos de 20 segundos. En lugar de ser cada segundo una entrada, se suman los valores de estas agrupaciones. Se recomienda valores pares en esta variable, pues se realizan agrupaciones solapadas, es decir, primero se comienza en el segundo 0 y se realizan agrupaciones de 20 segundos, y a continuación se vuelven a agrupar pero comenzando en el segundo 10 que es la mitad del intervalo de agrupación, en este caso 20 segundos.

9. KFolds, o número de particiones en la validación cruzada. Este valor durante el análisis mostrado se estableció en 10.

4.2.4. Resultados obtenidos

Los datos obtenidos se generan en gráficas que permiten su manipulación y análisis detallado. Todos los resultados que se muestran a continuación son generados mediante ficheros HTML. Este formato permite visualizar con más detalle las gráficas, mostrando solo las trazas que deseamos analizar. Además si el gráfico es de 3 dimensiones, permite rotar los resultados y realizar un análisis en profundidad. Debido a las limitaciones del formato PDF, se muestran varias imágenes con los gráficos generados con distintas perspectivas y algunos ejemplos.

Los resultados obtenidos tras la ejecución del algoritmo de predicción se muestran en la Figura 4.8 y Figura 4.9. Este gráfico muestra la tasa de acierto (eje z), para el número de problema mostrado en el eje x cuando el algoritmo de predicción ha sido entrenado con las visualizaciones del vídeo con el id mostrado en el eje y. Para representar en dos dimensiones el gráfico, se muestra desde la perspectiva superior (Figura 4.8). De este modo podemos ver el plano xy, Vídeo-Problema, y para representar los valores del eje z utilizamos una escala de colores, detallada en el gráfico.

Observando la Figura 4.8 podemos ver una diagonal donde los resultados de la predicción obtienen su valor máximo. Esta diagonal comienza al inicio del curso, y termina antes del examen. Es decir, se puede observar la relación temporal entre problemas y vídeos. La visualización de los primeros vídeos está directamente relacionada con la tasa de acierto al predecir los resultados de los problemas. Esta relación avanza temporalmente, y termina con los últimos vídeos del curso y antes de las preguntas del examen.

Si bien, durante el inicio del curso, hay gran cantidad de usuarios que visualizan los vídeos, pero que no tienen interés en el curso. Estos alumnos aunque, sí visualizan el vídeo, responden prácticamente al azar. Esto dificulta mucho al algoritmo de predicción encontrar patrones al inicio del curso, por lo que los valores de sus resultados en general, no solo en la diagonal, son más bajos. Esto puede observarse en la Figura 4.9, donde los valores medios en la tasa de acierto aumentan conforme avanza el curso.

Si observamos los problemas del examen en la Figura 4.8 parece que solo los últimos vídeos son importantes para acertar el resultado de los problemas. Esto podría implicar que en el examen la gran mayoría de preguntas están relacionadas con la parte final del temario. Si bien, es cierto que debido al contenido de este curso el temario del final engloba los conceptos del inicio, esta no es la razón que hace obtener mejores resultados a los últimos vídeos. Algunos alumnos, aun cuando no han visto todo el material del curso, realizan el examen, bien porque saben que no tienen la nota suficiente o simplemente por intentarlo. Existe una relación entre la continuidad en el curso, es decir, una relación entre visualizar los últimos vídeos, y acertar las respuestas del examen Por esta razón, los últimos vídeos obtienen mejores tasas de acierto en la predicción.

También podemos ver en la Figura 4.8 problemas donde la diferencia entre los resultados de distintos vídeos es mínima, a lo que anteriormente nos hemos referido con el nombre de "cresta". Estas "crestas" significaban que las visualizaciones de los distintos vídeo predicen resultados similares, no hay un vídeo cuya relación destaque sobre el resto. De este modo podemos revisar el problema y buscar si su contenido se explica en algún vídeo, y si debería reforzarse las explicaciones y el material en el que se basa la pregunta.

Obtenidos los datos de correlación entre vídeos y problemas hemos seleccionado aquellos que guardan mayor relación, pero es necesario poder visualizarlos. Para la representación de las relaciones hemos escogido una gráfica para cada vídeo, con todos los problemas relacionados. Esto supone que algunas gráficas no contengan apenas datos, y otras, como sucede en la de los últimos vídeos, contengan demasiados problemas. En estas graficas se muestra la visualización de los usuarios que aciertan el problema frente a quienes fallan.

Para obtener los valores de cada problema en el gráfico se suman las visualizaciones de los alumnos que acertaron la pregunta y se divide entre el número de alumnos que acertaron. Se suma también el total de aquellos que fallaron y se divide por el número de alumnos que fallaron. Estas sumas se realizan en cada segundo del vídeo, y la resta de estas dos cantidades es la que se muestra en la gráfica. De manera que si un problema alcanza el valor 1 en algún punto de la gráfica, significa que, de media, los usuarios que acertaron vieron 1 vez más ese segundo del vídeo. Mientras que si alcanza valores negativos, serán los usuarios que fallaron quienes vieron más esa parte del vídeo.

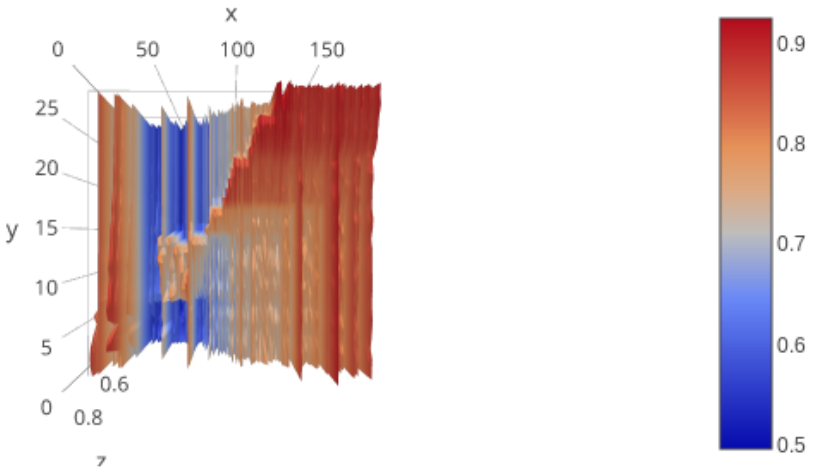

Figura 4.8: Representación de la relación entre Problemas y Vídeos

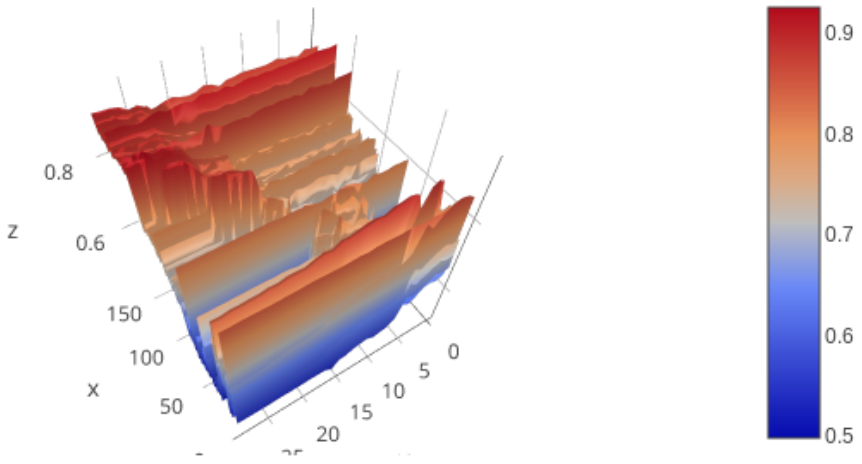

Figura 4.9: Representación de la relación entre Problemas y Vídeos en 3D

Como ya hemos mencionado la visualización del gráfico es interactiva. De forma automática la herramienta utilizada ajusta las escalas para maximizar el espacio visualizado, por ello es importante fijarse en la escala de los ejes. Distintas escalas no son adecuadas para una comparativa justa de los datos.

A continuación mostramos algunos ejemplos de las gráficas generadas y comentamos los detalles más relevantes.

En la Figura 4.10 vemos que el vídeo 15 tiene bastantes problemas relacionados. Fijándose en la escala, observamos que la diferencia es tan solo de 0.2 visualizaciones. Aunque puede parecer poco, pues los alumnos que aciertan el problema tan solo ven el vídeo de media 0.2 veces más, es una diferencia suficiente para que el algoritmo de predicción obtenga buenos resultados. Se puede analizar individualmente cada vídeo, ver qué valores son más altos, o incluso si existen valores negativos, pues supondría que fallaron la pregunta los más que vieron la parte correspondiente del vídeo.

En la Figura 4.11 se muestran los problemas para el vídeo 20. Se observa que hay tres problemas cuya diferencia entre aquellos que acertaron y fallaron es casi 0. Esto suele ser consecuencia de aquellos problemas que casi todos los vídeos obtenían la misma puntuación, había "crestas", pudiendo relacionar un vídeo y un problema que apenas guardan relación. Frente a estos problemas sin relación tenemos el problema 94, en el que podemos ver que aquellos usuarios que han visto el vídeo más veces suelen acertar más a menudo. Aunque con valores cercanos al 0.1, podemos ver que el problema 146 es más visto por aquellos usuarios que fallaron la pregunta. Dado que es un problema del examen y bastante alejado temporalmente del vídeo 20, esta relación no implica que el material del vídeo 20 sea incorrecto. Esta relación puede ser casualidad o implicar que su contenido puede provocar confusión en algunos usuarios.

El análisis de estos resultados siempre se debe contrastar con los datos del curso. Aunque la cantidad de datos es grande existe la posibilidad de tener un problema para el que, por ejemplo, la cantidad de alumnos que fallan sea mínima. La media de los usuarios que acertaron es válida, pero la de los fallos se basaría solo en unos cuantos alumnos. Por lo tanto, es importante comprender que disponemos de una herramienta que ayuda a la visualización y análisis de los datos, pero que estos deben corroborarse con los detalles particulares de cada vídeo y problema.

Finalmente, mostramos en la Figura 4.12 del vídeo 26 otro punto a analizar. En este caso los valores del eje y son cercanos a 1, por lo que la relación es clara. Para los problemas 144 y 145, si nos fijamos en los estudiantes que más aciertos consiguen, vemos que la visualización es casi idéntica durante todo el vídeo. Esto implica que los usuarios que acertaron el problema 145, son en gran parte aquellos que respondieron correctamente el problema 144.

Comparar entre problemas de un mismo vídeo es la razón por la cual se escogió esta representación. Por una parte en el capítulo 3 mostramos las visualizaciones de cada vídeo. Comparando la forma de la línea de cada problema en el gráfico, es posible analizar qué zonas, dentro del vídeo, son las que suponen la diferencia para cada problema. Por ejemplo, cerca del segundo 150, los problemas 144 y 145 adquieren gran importancia, mientras que en el resto de problemas los usuarios que visualizaron más veces esta parte no obtuvieron mejores resultados. En el resto de problemas no parece haber una zona que destaque, sino que es la visualización del vídeo en sí, la que ayuda a acertar la pregunta.

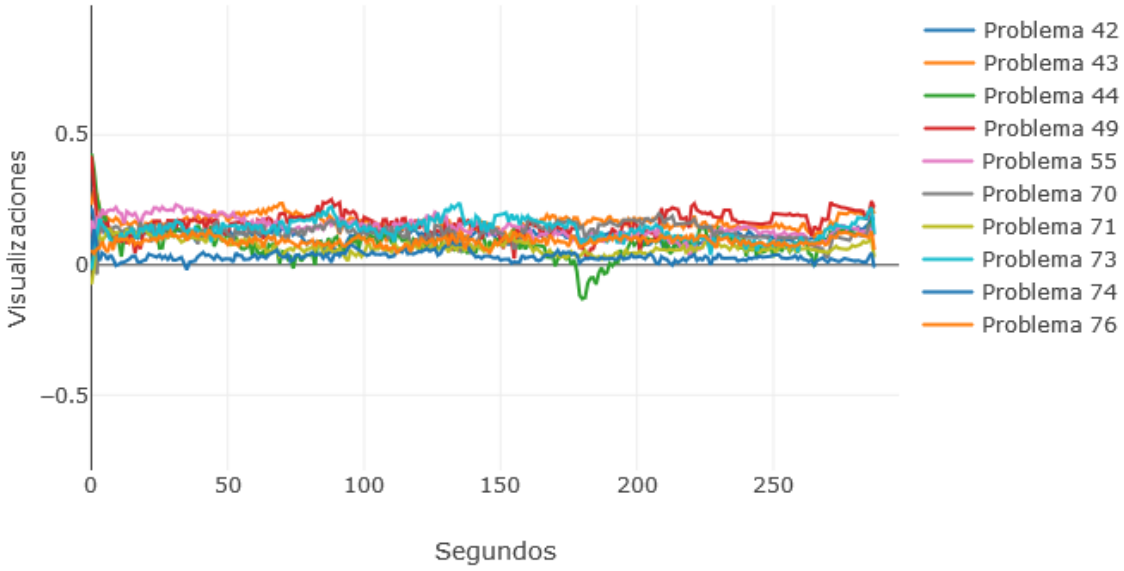

Figura 4.10: Relación de aciertos y fallos en el vídeo 15

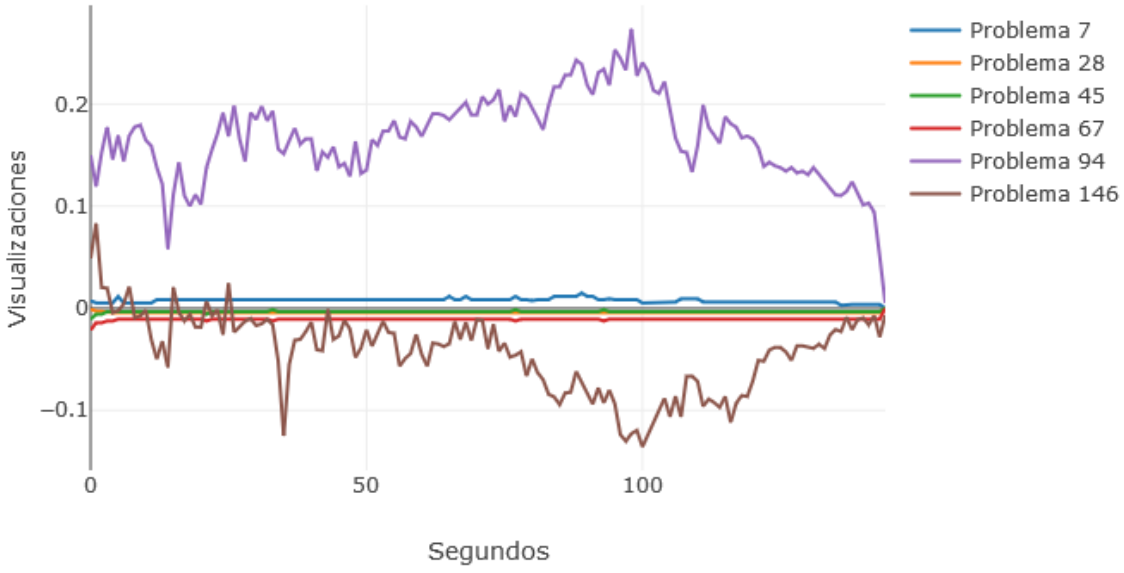

Figura 4.11: Relación de aciertos y fallos en el vídeo 20

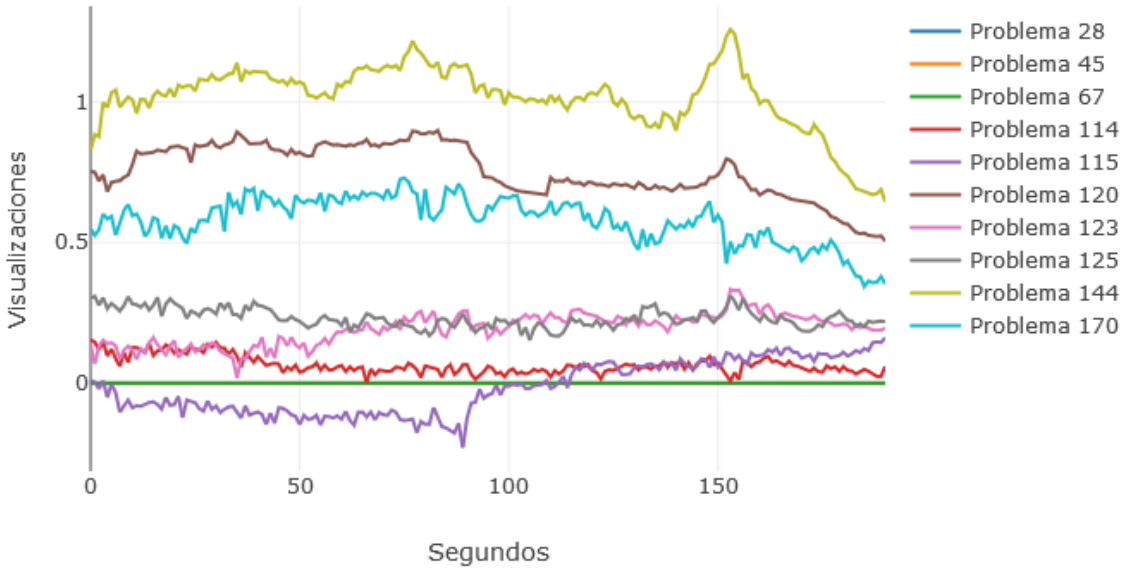

Figura 4.12: Relación de aciertos y fallos en el vídeo 26

5

Conclusiones y trabajo futuro

El uso de la tecnología en las aulas es el futuro. En este TFG nos hemos centrado en facilitar su uso, y sobre todo en ayudar a los docentes a mejorar la calidad de los cursos. Para ello hemos desarrollado una serie de herramientas que ayudan al análisis de los resultados y del curso en sí mismo. No obstante, el objetivo no ha sido procesar los datos del curso, y obtener resultados para ser almacenados sino que el esfuerzo se centra en detectar los problemas, ofrecer las herramientas para conocer los motivos y tratar de solucionarlos.

Existen diferentes herramientas, metodológicas y distinto material multimedia usado en cada uno de los cursos impartidos. Este proyecto se ha centrado en los dos elementos más básicos y comunes a casi todos los cursos: el uso de vídeos y de ejercicios. Hemos conseguido detectar qué partes de un vídeo son más visualizadas, si la visualización de los vídeos disminuye según este avanza y detectar qué zonas son responsables de acaparar la atención de los alumnos. Esto permite no solo detectar fallos en los vídeos, sino relacionarlos con los problemas e identificar los motivos.

Si somos capaces de detectar un problema en los vídeos, podemos solucionarlo, mejorando cada año la calidad del material. Incluso será posible detectar los problemas durante la impartición del curso y tratar de solucionarlos. El uso principal es detectar problemas generales en los vídeos, centrando los esfuerzos en reducir aquellos trozos de un vídeo que aumentan los fallos en los ejercicios y alargar las partes que ayudan a responder correctamente. Pero no solo podemos detectar problemas generales, también sería posible el uso de la herramienta con alumnos individualmente. Es decir, analizar cuáles son las razones que diferenciaron a los alumnos que acertaron de aquellos que fallaron, y sus semejanzas con aquellos que también fallaron.

Es más, no solo permite mejorar los vídeos, también conocer que partes del vídeo son indiferentes en los resultados. Ayudando al docente a detectar áreas del temario por las que no se les examina a los alumnos. O conocer si alguna pregunta es errónea o como de relacionadas están varias preguntas entre sí.

5.1. Trabajo futuro

Durante todo el proyecto las herramientas desarrolladas se han centrado en facilitar la visualización de los resultados pero sin perder de vista los datos reales. Una de las líneas de trabajo futuro será la automatización del análisis. Si bien se han utilizado distintos métodos para interpretar y entender los resultados obtenidos, éstos necesitan ser visualizados por una persona. Alojar en un servidor, con acceso a los datos del curso en tiempo real, un programa que periódicamente ejecutase el proyecto, facilitaría al profesorado analizar estos datos de forma temprana.

Incluso se podrán desarrollar herramientas que a partir de los datos generados permitan automatizar la detección de problemas y las causas. De este modo podría avisar al docente de forma temprana, evaluar la recomendación y aplicar las modificaciones necesarias.

Pero el potencial de esta herramienta no solo se centra en el curso y el profesorado, sino en poder ayudar a los alumnos. Si la detención de problemas se hace de forma periódica, podemos identificar qué material es el que distingue a los alumnos que aciertan de aquellos que fallan y recomendar individualmente a cada alumno reforzar su conocimiento en el área que se ha detectado el problema. Y no solo se trataría de avisar al alumno sino de indicarle que material debe consultar para poder aprobar.

Además de ayudar a los alumnos de forma individual se puede hacer en el curso en general. Se puede aportar material auxiliar que en un principio estuviera oculto, y que el programa incluiría de forma automática como material recomendado si se detectan problemas en ciertos temas.

Es este potencial en el que destaca esta herramienta sobre aquellas que se centran en el análisis sin intervención. Solo automatizando su se usó abre una gran variedad de oportunidades que podrían mejorar enormemente la calidad del material y la experiencia de los alumnos.

Bibliografía

[1] M. Raposo-Rivas, E. Martínez-Figueira, and J. Sarmiento Campos. Un estudio sobre los componentes pedagógicos de los cursos online masivos. pages 27–35, 2005.

[2] Lic. Mculha Martínez and Lic. Domingo Merlino. *Elementos de Matemática*, chapter Nuevas Tecnologías y Educación, pages 27–34. Universidad CAECE, 2001.

[3] Javier González Romero. Iii estudio sobre el uso de la tecnología en el aula. `https://www.realinfluencers.es/2017/06/19/estudio_tic_aula/`, 2017. [Online; accessed Jun 21, 2018].

[4] Bart Rienties, Simon Cross, and Zdenek Zdrahal. *Big Data and Learning Analytics in Higher Education*, chapter Chapter 10 Implementing a Learning Analytics Intervention and Evaluation Framework: What Works?, pages 147–166. Springer International Publishing Switzerland, 2017.

[5] Yannick Meier, Jie Xu, Onur Atan, and Mihaela van der Schaar Fellow. Predicting grades. *IEEE Transaction on signal processing, Vol. 64, No. 4*, 2016.

[6] Sunil Ray. Essentials of machine learning algorithms (with python and r codes). `https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/`, 2017. [Online; accessed Jun 21, 2018].

[7] Niall Sclater. Planning interventions with at-risk students — effective learning analytics. `https://analytics.jiscinvolve.org/wp/2017/06/26/planning-interventions-with-at-risk-students/`, 2017. [Online; accessed Jun 21, 2018].

[8] JISC (Joint Information Systems Committee). Learning analytics interventions - what's really going on? `https://accessibility.jiscinvolve.org/wp/2017/07/24/learning-analytics-interventions-whats-really-going/`, 2017. [Online; accessed Jun 21, 2018].

[9] Miguel Ángel González-Gallego Sosa. Predicción y análisis de interacciones de usuarios en plataformas de enseñanza online. *Trabajo Final de Grado, EPS (UAM)*, 2016.

Glosario de acrónimos

EPS	Escuela Politécnica Superior
JSON	JavaScript Object Notation
MOOC	Massive Online Course
TFG	Trabajo de fin de Grado
UAM	Universidad Autónoma de Madrid

Anexos

A

Documentos del curso

1. Semana 1
    a) 1.1. El Entorno de Desarrollo de Android
    b) 1.2a. Dispositivos virtuales
    c) 1.2b. El gestor del SDK
    d) 1.3. Dispositivos virtuales
    e) 1.4. La primera aplicación
    f) 1.5. Estructura de un proyecto
    g) 1.6. Una Introducción a XML
    h) 1.7. Recursos
2. Semana 2
    a) 2.1. La Primera Interfaz de usuario
    b) 2.2. Márgenes y espaciado
    c) 2.3. Gravedad
    d) 2.4. Pesos
    e) 2.5. RelativeLayout
    f) 2.6. Otros contenedores
    g) 2.7. Tamaños de pantalla y unidades de medida
    h) 2.8. Otras vistas
3. Semana 3
    a) 3.1. El Atributo android: onClick
    b) 3.2. Escuchadores de eventos
    c) 3.3. La lógica del juego

4. Semana 4
    a) 4.1. El ciclo de vida de una actividad
    b) 4.2. Música en tu Dispositivo
    c) 4.3. Orientación del Dispositivo
    d) 4.4. Intenciones
    e) 4.5. Arrancar Aplicaciones Integradas de Android
5. Semana 5
    a) 5.1. Fragmentos
    b) 5.2. Menús
    c) 5.3. Diálogos
    d) 5.4. Preferencias
6. Semana 6
    a) 6.1. Animaciones
    b) 6.2. Internacionalización
    c) 6.3. Publicación

B

Gráficas del análisis de los datos

A continuación, se muestran las gráficas de visualizaciones totales mostradas en el capítulo 4 con distintos valores para el filtrado de usuarios por eventos.

B.1. Número de visualizaciones totales

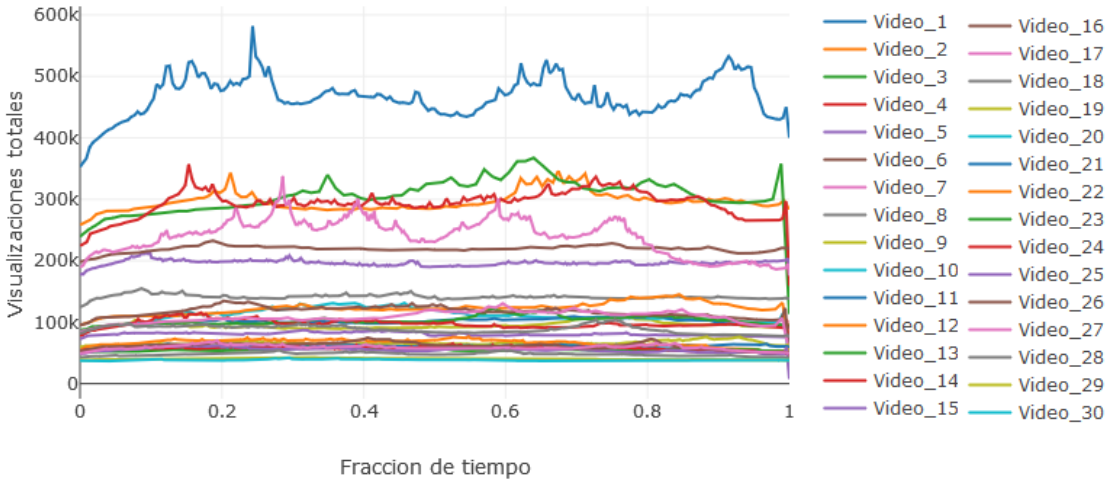

Figura B.1: Número visualizaciones totales sin filtro (Filtro 0)

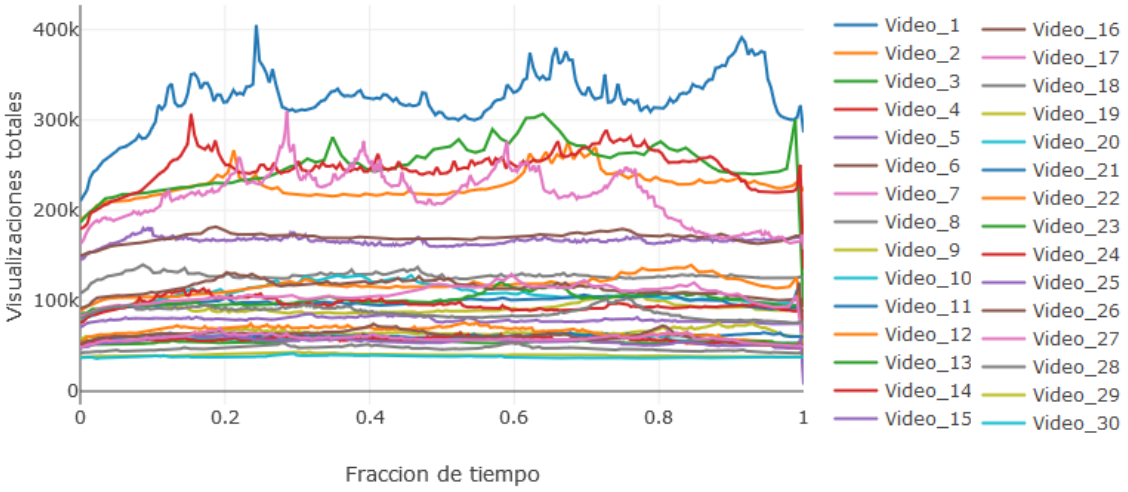

Figura B.2: Número visualizaciones totales con filtro 150

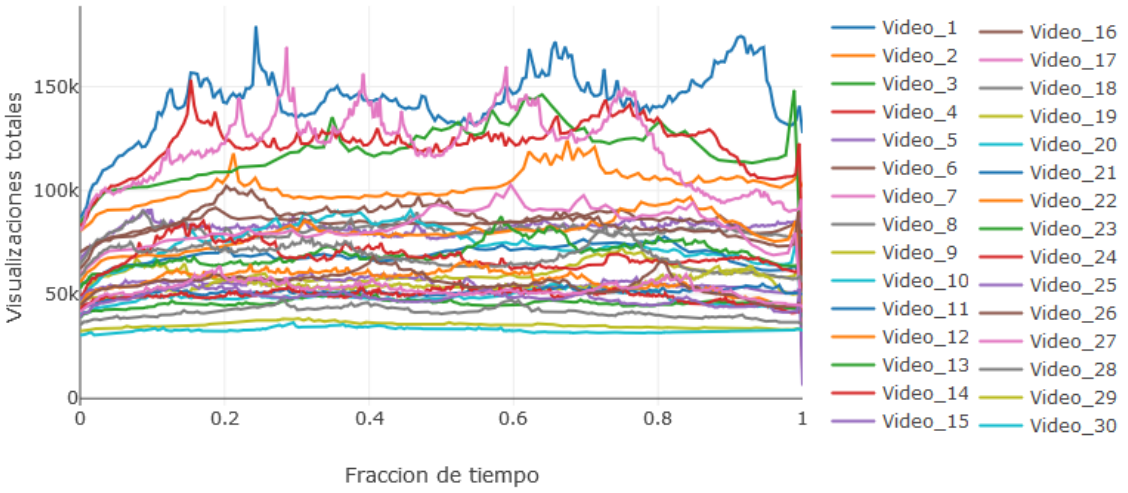

Figura B.3: Número visualizaciones totales con filtro 500

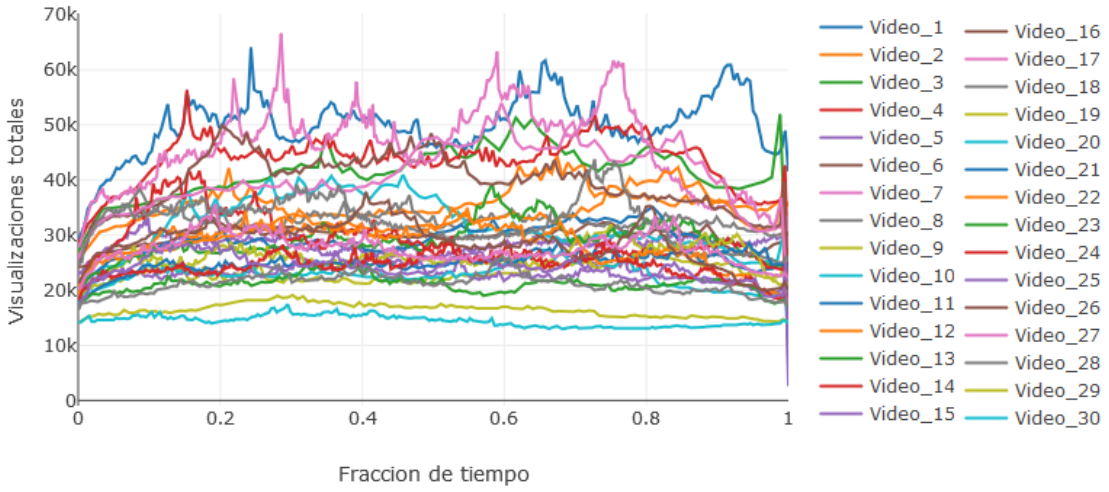

Figura B.4: Número visualizaciones totales con filtro 1000